

## NEISSERIAL ANTIGENS

This application is a continuation-in-part of international patent application PCT/IB98/01665, filed October 9, 1998, from which priority is claimed under 35 U.S.C. § 119.

This invention relates to antigens from *Neisseria* bacteria.

### 5 BACKGROUND ART

*Neisseria meningitidis* and *Neisseria gonorrhoeae* are non-motile, gram negative diplococci that are pathogenic in humans. *N.meningitidis* colonises the pharynx and causes meningitis (and, occasionally, septicaemia in the absence of meningitis); *N.gonorrhoeae* colonises the genital tract and causes gonorrhea. Although colonising different areas of the body and causing completely different diseases, the two pathogens are closely related, although one feature that clearly differentiates meningococcus from gonococcus is the presence of a polysaccharide capsule that is present in all pathogenic meningococci.

*N.gonorrhoeae* caused approximately 800,000 cases per year during the period 1983-1990 in the United States alone (chapter by Meitzner & Cohen, "Vaccines Against Gonococcal Infection", In: *New Generation Vaccines*, 2nd edition, ed. Levine, Woodrow, Kaper, & Cobon, Marcel Dekker, New York, 1997, pp.817-842). The disease causes significant morbidity but limited mortality. Vaccination against *N.gonorrhoeae* would be highly desirable, but repeated attempts have failed. The main candidate antigens for this vaccine are surface-exposed proteins such as pili, porins, opacity-associated proteins (Opas) and other surface-exposed proteins such as the Lip, Laz, IgA1 protease and transferrin-binding proteins. The lipooligosaccharide (LOS) has also been suggested as vaccine (Meitzner & Cohen, *supra*).

*N.meningitidis* causes both endemic and epidemic disease. In the United States the attack rate is 0.6-1 per 100,000 persons per year, and it can be much greater during outbreaks (see Lieberman *et al.* (1996) Safety and Immunogenicity of a Serogroups A/C *Neisseria meningitidis* Oligosaccharide-Protein Conjugate Vaccine in Young Children. *JAMA* 275(19):1499-1503; Schuchat *et al* (1997) Bacterial Meningitis in the United States in 1995. *N Engl J Med* 337(14):970-976). In developing countries, endemic disease rates are much higher and during epidemics incidence rates can reach 500 cases per 100,000 persons per year. Mortality is extremely high, at

10-20% in the United States, and much higher in developing countries. Following the introduction of the conjugate vaccine against *Haemophilus influenzae*, *N. meningitidis* is the major cause of bacterial meningitis at all ages in the United States (Schuchat *et al* (1997) *supra*).

Based on the organism's capsular polysaccharide, 12 serogroups of *N.meningitidis* have been identified. Group A is the pathogen most often implicated in epidemic disease in sub-Saharan Africa. Serogroups B and C are responsible for the vast majority of cases in the United States and in most developed countries. Serogroups W135 and Y are responsible for the rest of the cases in the United States and developed countries. The meningococcal vaccine currently in use is a tetravalent polysaccharide vaccine composed of serogroups A, C, Y and W135. Although efficacious in adolescents and adults, it induces a poor immune response and short duration of protection, and cannot be used in infants [*eg.* Morbidity and Mortality weekly report, Vol.46, No. RR-5 (1997)]. This is because polysaccharides are T-cell independent antigens that induce a weak immune response that cannot be boosted by repeated immunization. Following the success of the vaccination against *H.influenzae*, conjugate vaccines against serogroups A and C have been developed and are at the final stage of clinical testing (Zollinger WD "New and Improved Vaccines Against Meningococcal Disease" in: *New Generation Vaccines, supra*, pp. 469-488; Lieberman *et al* (1996) *supra*; Costantino *et al* (1992) Development and phase I clinical testing of a conjugate vaccine against meningococcus A and C. *Vaccine* 10:691-698).

Meningococcus B remains a problem, however. This serotype currently is responsible for approximately 50% of total meningitis in the United States, Europe, and South America. The polysaccharide approach cannot be used because the menB capsular polysaccharide is a polymer of  $\alpha(2-8)$ -linked *N*-acetyl neuraminic acid that is also present in mammalian tissue. This results in tolerance to the antigen; indeed, if an immune response were elicited, it would be anti-self, and therefore undesirable. In order to avoid induction of autoimmunity and to induce a protective immune response, the capsular polysaccharide has, for instance, been chemically modified substituting the *N*-acetyl groups with *N*-propionyl groups, leaving the specific antigenicity unaltered (Romero & Outschoorn (1994) Current status of Meningococcal group B vaccine candidates: capsular or non-capsular? *Clin Microbiol Rev* 7(4):559-575).

Alternative approaches to menB vaccines have used complex mixtures of outer membrane proteins (OMPs), containing either the OMPs alone, or OMPs enriched in porins, or deleted of the class 4



OMP's that are believed to induce antibodies that block bactericidal activity. This approach produces vaccines that are not well characterized. They are able to protect against the homologous strain, but are not effective at large where there are many antigenic variants of the outer membrane proteins. To overcome the antigenic variability, multivalent vaccines containing up to nine different porins have been constructed (eg. Poolman JT (1992) Development of a meningococcal vaccine. *Infect. Agents Dis.* 4:13-28). Additional proteins to be used in outer membrane vaccines have been the opa and opc proteins, but none of these approaches have been able to overcome the antigenic variability (eg. Ala'Aldeen & Borriello (1996) The meningococcal transferrin-binding proteins 1 and 2 are both surface exposed and generate bactericidal antibodies capable of killing homologous and heterologous strains. *Vaccine* 14(1):49-53).

A certain amount of sequence data is available for meningococcal and gonococcal genes and proteins (eg. EP-A-0467714, WO96/29412), but this is by no means complete. The provision of further sequences could provide an opportunity to identify secreted or surface-exposed proteins that are presumed targets for the immune system and which are not antigenically variable. For instance, some of the identified proteins could be components of efficacious vaccines against meningococcus B, some could be components of vaccines against all meningococcal serotypes, and others could be components of vaccines against all pathogenic *Neisseriae*.

### THE INVENTION

The invention provides proteins comprising the Neisserial amino acid sequences disclosed in the examples. These sequences relate to *N.meningitidis* or *N.gonorrhoeae*.

It also provides proteins comprising sequences homologous (*ie.* having sequence identity) to the Neisserial amino acid sequences disclosed in the examples. Depending on the particular sequence, the degree of identity is preferably greater than 50% (eg. 65%, 80%, 90%, or more). These homologous proteins include mutants and allelic variants of the sequences disclosed in the examples. Typically, 50% identity or more between two proteins is considered to be an indication of functional equivalence. Identity between the proteins is preferably determined by the Smith-Waterman homology search algorithm as implemented in the MPSRCH program (Oxford Molecular), using an affine gap search with parameters *gap open penalty*=12 and *gap extension penalty*=1.

The invention further provides proteins comprising fragments of the Neisserial amino acid sequences disclosed in the examples. The fragments should comprise at least  $n$  consecutive amino acids from the sequences and, depending on the particular sequence,  $n$  is 7 or more (eg. 8, 10, 12, 14, 16, 18, 20 or more). Preferably the fragments comprise an epitope from the sequence.

- 5 The proteins of the invention can, of course, be prepared by various means (eg. recombinant expression, purification from cell culture, chemical synthesis *etc.*) and in various forms (eg. native, fusions *etc.*). They are preferably prepared in substantially pure or isolated form (*ie.* substantially free from other Neisserial or host cell proteins)

- 10 According to a further aspect, the invention provides antibodies which bind to these proteins. These may be polyclonal or monoclonal and may be produced by any suitable means.

According to a further aspect, the invention provides nucleic acid comprising the Neisserial nucleotide sequences disclosed in the examples. In addition, the invention provides nucleic acid comprising sequences homologous (*ie.* having sequence identity) to the Neisserial nucleotide sequences disclosed in the examples.

- 15 Furthermore, the invention provides nucleic acid which can hybridise to the Neisserial nucleic acid disclosed in the examples, preferably under "high stringency" conditions (eg. 65°C in a 0.1xSSC, 0.5% SDS solution).

- 20 Nucleic acid comprising fragments of these sequences are also provided. These should comprise at least  $n$  consecutive nucleotides from the Neisserial sequences and, depending on the particular sequence,  $n$  is 10 or more (eg 12, 14, 15, 18, 20, 25, 30, 35, 40 or more).

According to a further aspect, the invention provides nucleic acid encoding the proteins and protein fragments of the invention.

It should also be appreciated that the invention provides nucleic acid comprising sequences complementary to those described above (eg. for antisense or probing purposes).

Nucleic acid according to the invention can, of course, be prepared in many ways (*eg.* by chemical synthesis, from genomic or cDNA libraries, from the organism itself *etc.*) and can take various forms (*eg.* single stranded, double stranded, vectors, probes *etc.*).

In addition, the term "nucleic acid" includes DNA and RNA, and also their analogues, such as those containing modified backbones, and also peptide nucleic acids (PNA) *etc.*

According to a further aspect, the invention provides vectors comprising nucleotide sequences of the invention (*eg.* expression vectors) and host cells transformed with such vectors.

According to a further aspect, the invention provides compositions comprising protein, antibody, and/or nucleic acid according to the invention. These compositions may be suitable as vaccines, for instance, or as diagnostic reagents, or as immunogenic compositions.

The invention also provides nucleic acid, protein, or antibody according to the invention for use as medicaments (*eg.* as vaccines) or as diagnostic reagents. It also provides the use of nucleic acid, protein, or antibody according to the invention in the manufacture of: (i) a medicament for treating or preventing infection due to Neisserial bacteria; (ii) a diagnostic reagent for detecting the presence of Neisserial bacteria or of antibodies raised against Neisserial bacteria; and/or (iii) a reagent which can raise antibodies against Neisserial bacteria. Said Neisserial bacteria may be any species or strain (such as *N.gonorrhoeae*, or any strain of *N.meningitidis*, such as strain A, strain B or strain C).

The invention also provides a method of treating a patient, comprising administering to the patient a therapeutically effective amount of nucleic acid, protein, and/or antibody according to the invention.

According to further aspects, the invention provides various processes.

A process for producing proteins of the invention is provided, comprising the step of culturing a host cell according to the invention under conditions which induce protein expression.

A process for producing protein or nucleic acid of the invention is provided, wherein the the protein or nucleic acid is synthesised in part or in whole using chemical means.

A process for detecting polynucleotides of the invention is provided, comprising the steps of: (a) contacting a nucleic probe according to the invention with a biological sample under hybridizing conditions to form duplexes; and (b) detecting said duplexes.

5 A process for detecting proteins of the invention is provided, comprising the steps of: (a) contacting an antibody according to the invention with a biological sample under conditions suitable for the formation of an antibody-antigen complexes; and (b) detecting said complexes.

A summary of standard techniques and procedures which may be employed in order to perform the invention (eg. to utilise the disclosed sequences for vaccination or diagnostic purposes) follows. This summary is not a limitation on the invention but, rather, gives examples that may be used, but  
10 are not required.

#### General

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature eg. Sambrook  
15 *Molecular Cloning: A Laboratory Manual, Second Edition* (1989); *DNA Cloning, Volumes I and ii* (D.N Glover ed. 1985); *Oligonucleotide Synthesis* (M.J. Gait ed, 1984); *Nucleic Acid Hybridization* (B.D. Hames & S.J. Higgins eds. 1984); *Transcription and Translation* (B.D. Hames & S.J. Higgins eds. 1984); *Animal Cell Culture* (R.I. Freshney ed. 1986); *Immobilized Cells and Enzymes* (IRL Press, 1986); B. Perbal, *A Practical Guide to Molecular Cloning* (1984); the  
20 *Methods in Enzymology* series (Academic Press, Inc.), especially volumes 154 & 155; *Gene Transfer Vectors for Mammalian Cells* (J.H. Miller and M.P. Calos eds. 1987, Cold Spring Harbor Laboratory); Mayer and Walker, eds. (1987), *Immunochemical Methods in Cell and Molecular Biology* (Academic Press, London); Scopes, (1987) *Protein Purification: Principles and Practice*, Second Edition (Springer-Verlag, N.Y.), and *Handbook of Experimental Immunology, Volumes*  
25 *I-IV* (D.M. Weir and C. C. Blackwell eds 1986).

Standard abbreviations for nucleotides and amino acids are used in this specification.

All publications, patents, and patent applications cited herein are incorporated in full by reference. In particular, the contents of UK patent applications 9723516.2, 9724190.5, 9724386.9, 9725158.1, 9726147.3, 9800759.4, and 9819016.8 are incorporated herein.

Definitions

A composition containing X is "substantially free of" Y when at least 85% by weight of the total X+Y in the composition is X. Preferably, X comprises at least about 90% by weight of the total of X+Y in the composition, more preferably at least about 95% or even 99% by weight.

- 5 The term "comprising" means "including" as well as "consisting" *eg.* a composition "comprising" X may consist exclusively of X or may include something additional to X, such as X+Y.

- 10 A "conserved" *Neisseria* amino acid fragment or protein is one that is present in a particular *Neisseria* protein in at least x% of *Neisseria*. The value of x may be 50% or more, *e.g.*, 66%, 75%, 80%, 90%, 95% or even 100% (*i.e.* the amino acid is found in the protein in question in all *Neisseria*). In order to determine whether an amino acid is "conserved" in a particular *Neisseria* protein, it is necessary to compare that amino acid residue in the sequences of the protein in question from a plurality of different *Neisseria* (a reference population). The reference population may include a number of different *Neisseria* species or may include a single species. The reference population may include a number of different serogroups of a particular species or a single serogroup. A preferred reference population consists of the 5 most common *Neisseria*. The term "heterologous" refers to two biological components that are not found together in nature. The components may be host cells, genes, or regulatory regions, such as promoters. Although the heterologous components are not found together in nature, they can function together, as when a promoter heterologous to a gene is operably linked to the gene. Another example is where a *Neisseria* sequence is heterologous to a mouse host cell. A further examples would be two epitopes from the same or different proteins which have been assembled in a single protein in an arrangement not found in nature.
- 20

- 25 An "origin of replication" is a polynucleotide sequence that initiates and regulates replication of polynucleotides, such as an expression vector. The origin of replication behaves as an autonomous unit of polynucleotide replication within a cell, capable of replication under its own control. An origin of replication may be needed for a vector to replicate in a particular host cell. With certain origins of replication, an expression vector can be reproduced at a high copy number in the presence of the appropriate proteins within the cell. Examples of origins are the autonomously replicating sequences, which are effective in yeast; and the viral T-antigen, effective in COS-7 cells.
- 30

A "mutant" sequence is defined as DNA, RNA or amino acid sequence differing from but having sequence identity with the native or disclosed sequence. Depending on the particular sequence, the degree of sequence identity between the native or disclosed sequence and the mutant sequence is preferably greater than 50% (eg. 60%, 70%, 80%, 90%, 95%, 99% or more, calculated using the Smith-Waterman algorithm as described above). As used herein, an "allelic variant" of a nucleic acid molecule, or region, for which nucleic acid sequence is provided herein is a nucleic acid molecule, or region, that occurs essentially at the same locus in the genome of another or second isolate, and that, due to natural variation caused by, for example, mutation or recombination, has a similar but not identical nucleic acid sequence. A coding region allelic variant typically encodes a protein having similar activity to that of the protein encoded by the gene to which it is being compared. An allelic variant can also comprise an alteration in the 5' or 3' untranslated regions of the gene, such as in regulatory control regions (eg. see US patent 5,753,235).

#### Expression systems

The Neisserial nucleotide sequences can be expressed in a variety of different expression systems; for example those used with mammalian cells, baculoviruses, plants, bacteria, and yeast.

##### i. Mammalian Systems

Mammalian expression systems are known in the art. A mammalian promoter is any DNA sequence capable of binding mammalian RNA polymerase and initiating the downstream (3') transcription of a coding sequence (eg. structural gene) into mRNA. A promoter will have a transcription initiating region, which is usually placed proximal to the 5' end of the coding sequence, and a TATA box, usually located 25-30 base pairs (bp) upstream of the transcription initiation site. The TATA box is thought to direct RNA polymerase II to begin RNA synthesis at the correct site. A mammalian promoter will also contain an upstream promoter element, usually located within 100 to 200 bp upstream of the TATA box. An upstream promoter element determines the rate at which transcription is initiated and can act in either orientation [Sambrook et al. (1989) "Expression of Cloned Genes in Mammalian Cells." In *Molecular Cloning: A Laboratory Manual*, 2nd ed.].

Mammalian viral genes are often highly expressed and have a broad host range; therefore sequences encoding mammalian viral genes provide particularly useful promoter sequences. Examples include the SV40 early promoter, mouse mammary tumor virus LTR promoter, adenovirus major late

promoter (Ad MLP), and herpes simplex virus promoter. In addition, sequences derived from non-viral genes, such as the murine metallothionein gene, also provide useful promoter sequences. Expression may be either constitutive or regulated (inducible), depending on the promoter can be induced with glucocorticoid in hormone-responsive cells.

- 5 The presence of an enhancer element (enhancer), combined with the promoter elements described above, will usually increase expression levels. An enhancer is a regulatory DNA sequence that can stimulate transcription up to 1000-fold when linked to homologous or heterologous promoters, with synthesis beginning at the normal RNA start site. Enhancers are also active when they are placed upstream or downstream from the transcription initiation site, in either normal or flipped orientation, or at a distance of more than 1000 nucleotides from the promoter [Maniatis et al. (1987) *Science* 236:1237; Alberts et al. (1989) *Molecular Biology of the Cell*, 2nd ed.]. Enhancer elements derived from viruses may be particularly useful, because they usually have a broader host range. Examples include the SV40 early gene enhancer [Dijkema et al (1985) *EMBO J.* 4:761] and the enhancer/promoters derived from the long terminal repeat (LTR) of the Rous Sarcoma Virus [Gorman et al. (1982b) *Proc. Natl. Acad. Sci.* 79:6777] and from human cytomegalovirus [Boshart et al. (1985) *Cell* 41:521]. Additionally, some enhancers are regulatable and become active only in the presence of an inducer, such as a hormone or metal ion [Sassone-Corsi and Borelli (1986) *Trends Genet.* 2:215; Maniatis et al. (1987) *Science* 236:1237].

- 20 A DNA molecule may be expressed intracellularly in mammalian cells. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein will always be a methionine, which is encoded by the ATG start codon. If desired, the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide.

- 25 Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provides for secretion of the foreign protein in mammalian cells. Preferably, there are processing sites encoded between the leader fragment and the foreign gene that can be cleaved either *in vivo* or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The adenovirus tripartite leader is an example of a leader sequence that provides for secretion of a foreign protein in mammalian cells.
- 30

Usually, transcription termination and polyadenylation sequences recognized by mammalian cells are regulatory regions located 3' to the translation stop codon and thus, together with the promoter elements, flank the coding sequence. The 3' terminus of the mature mRNA is formed by site-specific post-transcriptional cleavage and polyadenylation [Birnstiel et al. (1985) *Cell* 41:349; Proudfoot and Whitelaw (1988) "Termination and 3' end processing of eukaryotic RNA. In *Transcription and splicing* (ed. B.D. Hames and D.M. Glover); Proudfoot (1989) *Trends Biochem. Sci.* 14:105]. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminator/polyadenylation signals include those derived from SV40 [Sambrook et al (1989) "Expression of cloned genes in cultured mammalian cells." In *Molecular Cloning: A Laboratory Manual*].

Usually, the above described components, comprising a promoter, polyadenylation signal, and transcription termination sequence are put together into expression constructs. Enhancers, introns with functional splice donor and acceptor sites, and leader sequences may also be included in an expression construct, if desired. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (eg. plasmids) capable of stable maintenance in a host, such as mammalian cells or bacteria. Mammalian replication systems include those derived from animal viruses, which require trans-acting factors to replicate. For example, plasmids containing the replication systems of papovaviruses, such as SV40 [Gluzman (1981) *Cell* 23:175] or polyomavirus, replicate to extremely high copy number in the presence of the appropriate viral T antigen. Additional examples of mammalian replicons include those derived from bovine papillomavirus and Epstein-Barr virus. Additionally, the replicon may have two replicaton systems, thus allowing it to be maintained, for example, in mammalian cells for expression and in a prokaryotic host for cloning and amplification. Examples of such mammalian-bacteria shuttle vectors include pMT2 [Kaufman et al. (1989) *Mol. Cell. Biol.* 9:946] and pHEBO [Shimizu et al. (1986) *Mol. Cell. Biol.* 6:1074].

The transformation procedure used depends upon the host to be transformed. Methods for introduction of heterologous polynucleotides into mammalian cells are known in the art and include dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei.



Mammalian cell lines available as hosts for expression are known in the art and include many immortalized cell lines available from the American Type Culture Collection (ATCC), including but not limited to, Chinese hamster ovary (CHO) cells, HeLa cells, baby hamster kidney (BHK) cells, monkey kidney cells (COS), human hepatocellular carcinoma cells (eg. Hep G2), and a number of other cell lines.

## ii. Baculovirus Systems

The polynucleotide encoding the protein can also be inserted into a suitable insect expression vector, and is operably linked to the control elements within that vector. Vector construction employs techniques which are known in the art. Generally, the components of the expression system include a transfer vector, usually a bacterial plasmid, which contains both a fragment of the baculovirus genome, and a convenient restriction site for insertion of the heterologous gene or genes to be expressed; a wild type baculovirus with a sequence homologous to the baculovirus-specific fragment in the transfer vector (this allows for the homologous recombination of the heterologous gene in to the baculovirus genome); and appropriate insect host cells and growth media.

After inserting the DNA sequence encoding the protein into the transfer vector, the vector and the wild type viral genome are transfected into an insect host cell where the vector and viral genome are allowed to recombine. The packaged recombinant virus is expressed and recombinant plaques are identified and purified. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, *inter alia*, Invitrogen, San Diego CA ("MaxBac" kit). These techniques are generally known to those skilled in the art and fully described in Summers and Smith, *Texas Agricultural Experiment Station Bulletin No. 1555* (1987) (hereinafter "Summers and Smith").

Prior to inserting the DNA sequence encoding the protein into the baculovirus genome, the above described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are usually assembled into an intermediate transplacement construct (transfer vector). This construct may contain a single gene and operably linked regulatory elements; multiple genes, each with its owned set of operably linked regulatory elements; or multiple genes, regulated by the same set of regulatory elements. Intermediate transplacement constructs are often maintained in a replicon, such as an extrachromosomal element (eg. plasmids)

capable of stable maintenance in a host, such as a bacterium. The replicon will have a replication system, thus allowing it to be maintained in a suitable host for cloning and amplification.

Currently, the most commonly used transfer vector for introducing foreign genes into AcNPV is pAc373. Many other vectors, known to those of skill in the art, have also been designed. These  
5 include, for example, pVL985 (which alters the polyhedrin start codon from ATG to ATT, and which introduces a BamHI cloning site 32 basepairs downstream from the ATT; see Luckow and Summers, *Virology* (1989) 17:31.

The plasmid usually also contains the polyhedrin polyadenylation signal (Miller et al. (1988) *Ann. Rev. Microbiol.*, 42:177) and a prokaryotic ampicillin-resistance (*amp*) gene and origin of  
10 replication for selection and propagation in *E. coli*.

Baculovirus transfer vectors usually contain a baculovirus promoter. A baculovirus promoter is any DNA sequence capable of binding a baculovirus RNA polymerase and initiating the downstream (5' to 3') transcription of a coding sequence (eg. structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding  
15 sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A baculovirus transfer vector may also have a second domain called an enhancer, which, if present, is usually distal to the structural gene. Expression may be either regulated or constitutive.

Structural genes, abundantly transcribed at late times in a viral infection cycle, provide particularly  
20 useful promoter sequences. Examples include sequences derived from the gene encoding the viral polyhedron protein, Friesen et al., (1986) "The Regulation of Baculovirus Gene Expression," in: *The Molecular Biology of Baculoviruses* (ed. Walter Doerfler); EPO Publ. Nos. 127 839 and 155 476; and the gene encoding the p10 protein, Vlak et al., (1988), *J. Gen. Virol.* 69:765.

DNA encoding suitable signal sequences can be derived from genes for secreted insect or  
25 baculovirus proteins, such as the baculovirus polyhedrin gene (Carbonell et al. (1988) *Gene*, 73:409). Alternatively, since the signals for mammalian cell posttranslational modifications (such as signal peptide cleavage, proteolytic cleavage, and phosphorylation) appear to be recognized by insect cells, and the signals required for secretion and nuclear accumulation also appear to be conserved between the invertebrate cells and vertebrate cells, leaders of non-insect origin, such as

those derived from genes encoding human  $\alpha$ -interferon, Maeda et al., (1985), *Nature* 315:592; human gastrin-releasing peptide, Lebacqz-Verheyden et al., (1988), *Molec. Cell. Biol.* 8:3129; human IL-2, Smith et al., (1985) *Proc. Nat'l Acad. Sci. USA*, 82:8404; mouse IL-3, (Miyajima et al., (1987) *Gene* 58:273; and human glucocerebrosidase, Martin et al. (1988) *DNA*, 7:99, can also  
5 be used to provide for secretion in insects.

A recombinant polypeptide or polyprotein may be expressed intracellularly or, if it is expressed with the proper regulatory sequences, it can be secreted. Good intracellular expression of nonfused foreign proteins usually requires heterologous genes that ideally have a short leader sequence containing suitable translation initiation signals preceding an ATG start signal. If desired,  
10 methionine at the N-terminus may be cleaved from the mature protein by *in vitro* incubation with cyanogen bromide.

Alternatively, recombinant polyproteins or proteins which are not naturally secreted can be secreted from the insect cell by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provides for secretion of the foreign protein in insects. The  
15 leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the translocation of the protein into the endoplasmic reticulum.

After insertion of the DNA sequence and/or the gene encoding the expression product precursor of the protein, an insect cell host is co-transformed with the heterologous DNA of the transfer vector and the genomic DNA of wild type baculovirus -- usually by co-transfection. The promoter  
20 and transcription termination sequence of the construct will usually comprise a 2-5kb section of the baculovirus genome. Methods for introducing heterologous DNA into the desired site in the baculovirus virus are known in the art. (See Summers and Smith *supra*; Ju et al. (1987); Smith et al., *Mol. Cell. Biol.* (1983) 3:2156; and Luckow and Summers (1989)). For example, the insertion can be into a gene such as the polyhedrin gene, by homologous double crossover recombination;  
25 insertion can also be into a restriction enzyme site engineered into the desired baculovirus gene. Miller et al., (1989), *Bioessays* 4:91. The DNA sequence, when cloned in place of the polyhedrin gene in the expression vector, is flanked both 5' and 3' by polyhedrin-specific sequences and is positioned downstream of the polyhedrin promoter.

The newly formed baculovirus expression vector is subsequently packaged into an infectious recombinant baculovirus. Homologous recombination occurs at low frequency (between about 1% and about 5%); thus, the majority of the virus produced after cotransfection is still wild-type virus. Therefore, a method is necessary to identify recombinant viruses. An advantage of the expression system is a visual screen allowing recombinant viruses to be distinguished. The polyhedrin protein, which is produced by the native virus, is produced at very high levels in the nuclei of infected cells at late times after viral infection. Accumulated polyhedrin protein forms occlusion bodies that also contain embedded particles. These occlusion bodies, up to 15  $\mu\text{m}$  in size, are highly refractile, giving them a bright shiny appearance that is readily visualized under the light microscope. Cells infected with recombinant viruses lack occlusion bodies. To distinguish recombinant virus from wild-type virus, the transfection supernatant is plaqued onto a monolayer of insect cells by techniques known to those skilled in the art. Namely, the plaques are screened under the light microscope for the presence (indicative of wild-type virus) or absence (indicative of recombinant virus) of occlusion bodies. "Current Protocols in Microbiology" Vol. 2 (Ausubel et al. eds) at 16.8 (Supp. 10, 1990); Summers and Smith, *supra*; Miller et al. (1989).

Recombinant baculovirus expression vectors have been developed for infection into several insect cells. For example, recombinant baculoviruses have been developed for, *inter alia*: *Aedes aegypti*, *Autographa californica*, *Bombyx mori*, *Drosophila melanogaster*, *Spodoptera frugiperda*, and *Trichoplusia ni* (WO 89/046699; Carbonell et al., (1985) *J. Virol.* 56:153; Wright (1986) *Nature* 321:718; Smith et al., (1983) *Mol. Cell. Biol.* 3:2156; and see generally, Fraser, et al. (1989) *In Vitro Cell. Dev. Biol.* 25:225).

Cells and cell culture media are commercially available for both direct and fusion expression of heterologous polypeptides in a baculovirus/expression system; cell culture technology is generally known to those skilled in the art. See, eg. Summers and Smith *supra*.

The modified insect cells may then be grown in an appropriate nutrient medium, which allows for stable maintenance of the plasmid(s) present in the modified insect host. Where the expression product gene is under inducible control, the host may be grown to high density, and expression induced. Alternatively, where expression is constitutive, the product will be continuously expressed into the medium and the nutrient medium must be continuously circulated, while removing the product of interest and augmenting depleted nutrients. The product may be purified by such

techniques as chromatography, eg. HPLC, affinity chromatography, ion exchange chromatography, etc.; electrophoresis; density gradient centrifugation; solvent extraction, or the like. As appropriate, the product may be further purified, as required, so as to remove substantially any insect proteins which are also secreted in the medium or result from lysis of insect cells, so as to provide a product which is at least substantially free of host debris, eg. proteins, lipids and polysaccharides.

In order to obtain protein expression, recombinant host cells derived from the transformants are incubated under conditions which allow expression of the recombinant protein encoding sequence. These conditions will vary, dependent upon the host cell selected. However, the conditions are readily ascertainable to those of ordinary skill in the art, based upon what is known in the art.

### iii. Plant Systems

There are many plant cell culture and whole plant genetic expression systems known in the art. Exemplary plant cellular genetic expression systems include those described in patents, such as: US 5,693,506; US 5,659,122; and US 5,608,143. Additional examples of genetic expression in plant cell culture has been described by Zenk, *Phytochemistry* 30:3861-3863 (1991). Descriptions of plant protein signal peptides may be found in addition to the references described above in Vaulcombe et al., *Mol. Gen. Genet.* 209:33-40 (1987); Chandler et al., *Plant Molecular Biology* 3:407-418 (1984); Rogers, *J. Biol. Chem.* 260:3731-3738 (1985); Rothstein et al., *Gene* 55:353-356 (1987); Whittier et al., *Nucleic Acids Research* 15:2515-2535 (1987); Wirsal et al., *Molecular Microbiology* 3:3-14 (1989); Yu et al., *Gene* 122:247-253 (1992). A description of the regulation of plant gene expression by the phytohormone, gibberellic acid and secreted enzymes induced by gibberellic acid can be found in R.L. Jones and J. MacMillin, *Gibberellins*: in: *Advanced Plant Physiology*, Malcolm B. Wilkins, ed., 1984 Pitman Publishing Limited, London, pp. 21-52. References that describe other metabolically-regulated genes: Sheen, *Plant Cell*, 2:1027-1038(1990); Maas et al., *EMBO J.* 9:3447-3452 (1990); Benkel and Hickey, *Proc. Natl. Acad. Sci.* 84:1337-1339 (1987)

Typically, using techniques known in the art, a desired polynucleotide sequence is inserted into an expression cassette comprising genetic regulatory elements designed for operation in plants. The expression cassette is inserted into a desired expression vector with companion sequences upstream and downstream from the expression cassette suitable for expression in a plant host. The companion sequences will be of plasmid or viral origin and provide necessary characteristics to the

vector to permit the vectors to move DNA from an original cloning host, such as bacteria, to the desired plant host. The basic bacterial/plant vector construct will preferably provide a broad host range prokaryote replication origin; a prokaryote selectable marker; and, for *Agrobacterium* transformations, T DNA sequences for *Agrobacterium*-mediated transfer to plant chromosomes.

- 5 Where the heterologous gene is not readily amenable to detection, the construct will preferably also have a selectable marker gene suitable for determining if a plant cell has been transformed. A general review of suitable markers, for example for the members of the grass family, is found in Wilmink and Dons, 1993, *Plant Mol. Biol. Repr.*, 11(2):165-185.

- 10 Sequences suitable for permitting integration of the heterologous sequence into the plant genome are also recommended. These might include transposon sequences and the like for homologous recombination as well as Ti sequences which permit random insertion of a heterologous expression cassette into a plant genome. Suitable prokaryote selectable markers include resistance toward antibiotics such as ampicillin or tetracycline. Other DNA sequences encoding additional functions may also be present in the vector, as is known in the art.

- 15 The nucleic acid molecules of the subject invention may be included into an expression cassette for expression of the protein(s) of interest. Usually, there will be only one expression cassette, although two or more are feasible. The recombinant expression cassette will contain in addition to the heterologous protein encoding sequence the following elements, a promoter region, plant 5' untranslated sequences, initiation codon depending upon whether or not the structural gene comes  
20 equipped with one, and a transcription and translation termination sequence. Unique restriction enzyme sites at the 5' and 3' ends of the cassette allow for easy insertion into a pre-existing vector.

- A heterologous coding sequence may be for any protein relating to the present invention. The sequence encoding the protein of interest will encode a signal peptide which allows processing and translocation of the protein, as appropriate, and will usually lack any sequence which might result  
25 in the binding of the desired protein of the invention to a membrane. Since, for the most part, the transcriptional initiation region will be for a gene which is expressed and translocated during germination, by employing the signal peptide which provides for translocation, one may also provide for translocation of the protein of interest. In this way, the protein(s) of interest will be translocated from the cells in which they are expressed and may be efficiently harvested. Typically  
30 secretion in seeds are across the aleurone or scutellar epithelium layer into the endosperm of the

seed. While it is not required that the protein be secreted from the cells in which the protein is produced, this facilitates the isolation and purification of the recombinant protein.

Since the ultimate expression of the desired gene product will be in a eucaryotic cell it is desirable to determine whether any portion of the cloned gene contains sequences which will be processed out as introns by the host's splicosome machinery. If so, site-directed mutagenesis of the "intron" region may be conducted to prevent losing a portion of the genetic message as a false intron code, Reed and Maniatis, *Cell* 41:95-105, 1985.

The vector can be microinjected directly into plant cells by use of micropipettes to mechanically transfer the recombinant DNA. Crossway, *Mol. Gen. Genet.*, 202:179-185, 1985. The genetic material may also be transferred into the plant cell by using polyethylene glycol, Krens, et al., *Nature*, 296, 72-74, 1982. Another method of introduction of nucleic acid segments is high velocity ballistic penetration by small particles with the nucleic acid either within the matrix of small beads or particles, or on the surface, Klein, et al., *Nature*, 327, 70-73, 1987 and Knudsen and Muller, 1991, *Planta*, 185:330-336 teaching particle bombardment of barley endosperm to create transgenic barley. Yet another method of introduction would be fusion of protoplasts with other entities, either minicells, cells, lysosomes or other fusible lipid-surfaced bodies, Fraley, et al., *Proc. Natl. Acad. Sci. USA*, 79, 1859-1863, 1982.

The vector may also be introduced into the plant cells by electroporation. (Fromm et al., *Proc. Natl. Acad. Sci. USA* 82:5824, 1985). In this technique, plant protoplasts are electroporated in the presence of plasmids containing the gene construct. Electrical impulses of high field strength reversibly permeabilize biomembranes allowing the introduction of the plasmids. Electroporated plant protoplasts reform the cell wall, divide, and form plant callus.

All plants from which protoplasts can be isolated and cultured to give whole regenerated plants can be transformed by the present invention so that whole plants are recovered which contain the transferred gene. It is known that practically all plants can be regenerated from cultured cells or tissues, including but not limited to all major species of sugarcane, sugar beet, cotton, fruit and other trees, legumes and vegetables. Some suitable plants include, for example, species from the genera *Fragaria*, *Lotus*, *Medicago*, *Onobrychis*, *Trifolium*, *Trigonella*, *Vigna*, *Citrus*, *Linum*, *Geranium*, *Manihot*, *Daucus*, *Arabidopsis*, *Brassica*, *Raphanus*, *Sinapis*, *Atropa*, *Capsicum*, *Datura*, *Hyoscyamus*, *Lycopersion*, *Nicotiana*, *Solanum*, *Petunia*, *Digitalis*, *Majorana*, *Cichorium*,

*Helianthus, Lactuca, Bromus, Asparagus, Antirrhinum, Hererocallis, Nemesia, Pelargonium, Panicum, Pennisetum, Ranunculus, Senecio, Salpiglossis, Cucumis, Browaalia, Glycine, Lolium, Zea, Triticum, Sorghum, and Datura.*

Means for regeneration vary from species to species of plants, but generally a suspension of transformed protoplasts containing copies of the heterologous gene is first provided. Callus tissue is formed and shoots may be induced from callus and subsequently rooted. Alternatively, embryo formation can be induced from the protoplast suspension. These embryos germinate as natural embryos to form plants. The culture media will generally contain various amino acids and hormones, such as auxin and cytokinins. It is also advantageous to add glutamic acid and proline to the medium, especially for such species as corn and alfalfa. Shoots and roots normally develop simultaneously. Efficient regeneration will depend on the medium, on the genotype, and on the history of the culture. If these three variables are controlled, then regeneration is fully reproducible and repeatable.

In some plant cell culture systems, the desired protein of the invention may be excreted or alternatively, the protein may be extracted from the whole plant. Where the desired protein of the invention is secreted into the medium, it may be collected. Alternatively, the embryos and embryoless-half seeds or other plant tissue may be mechanically disrupted to release any secreted protein between cells and tissues. The mixture may be suspended in a buffer solution to retrieve soluble proteins. Conventional protein isolation and purification methods will be then used to purify the recombinant protein. Parameters of time, temperature pH, oxygen, and volumes will be adjusted through routine methods to optimize expression and recovery of heterologous protein.

#### iv. Bacterial Systems

Bacterial expression techniques are known in the art. A bacterial promoter is any DNA sequence capable of binding bacterial RNA polymerase and initiating the downstream (3') transcription of a coding sequence (eg. structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A bacterial promoter may also have a second domain called an operator, that may overlap an adjacent RNA polymerase binding site at which RNA synthesis begins. The operator permits negative regulated (inducible) transcription, as a gene repressor protein may bind the operator and



thereby inhibit transcription of a specific gene. Constitutive expression may occur in the absence of negative regulatory elements, such as the operator. In addition, positive regulation may be achieved by a gene activator protein binding sequence, which, if present is usually proximal (5') to the RNA polymerase binding sequence. An example of a gene activator protein is the catabolite activator protein (CAP), which helps initiate transcription of the lac operon in *Escherichia coli* (E. coli) [Raibaud *et al.* (1984) *Annu. Rev. Genet.* 18:173]. Regulated expression may therefore be either positive or negative, thereby either enhancing or reducing transcription.

Sequences encoding metabolic pathway enzymes provide particularly useful promoter sequences. Examples include promoter sequences derived from sugar metabolizing enzymes, such as galactose, lactose (*lac*) [Chang *et al.* (1977) *Nature* 198:1056], and maltose. Additional examples include promoter sequences derived from biosynthetic enzymes such as tryptophan (*trp*) [Goeddel *et al.* (1980) *Nuc. Acids Res.* 8:4057; Yelverton *et al.* (1981) *Nucl. Acids Res.* 9:731; US patent 4,738,921; EP-A-0036776 and EP-A-0121775]. The g-lactamase (*bla*) promoter system [Weissmann (1981) "The cloning of interferon and other mistakes." In *Interferon 3* (ed. I. Gresser)], bacteriophage lambda PL [Shimatake *et al.* (1981) *Nature* 292:128] and T5 [US patent 4,689,406] promoter systems also provide useful promoter sequences.

In addition, synthetic promoters which do not occur in nature also function as bacterial promoters. For example, transcription activation sequences of one bacterial or bacteriophage promoter may be joined with the operon sequences of another bacterial or bacteriophage promoter, creating a synthetic hybrid promoter [US patent 4,551,433]. For example, the *tac* promoter is a hybrid *trp-lac* promoter comprised of both *trp* promoter and *lac* operon sequences that is regulated by the *lac* repressor [Amann *et al.* (1983) *Gene* 25:167; de Boer *et al.* (1983) *Proc. Natl. Acad. Sci.* 80:21]. Furthermore, a bacterial promoter can include naturally occurring promoters of non-bacterial origin that have the ability to bind bacterial RNA polymerase and initiate transcription. A naturally occurring promoter of non-bacterial origin can also be coupled with a compatible RNA polymerase to produce high levels of expression of some genes in prokaryotes. The bacteriophage T7 RNA polymerase/promoter system is an example of a coupled promoter system [Studier *et al.* (1986) *J. Mol. Biol.* 189:113; Tabor *et al.* (1985) *Proc Natl. Acad. Sci.* 82:1074]. In addition, a hybrid promoter can also be comprised of a bacteriophage promoter and an *E. coli* operator region (EPO-A-0 267 851).

In addition to a functioning promoter sequence, an efficient ribosome binding site is also useful for the expression of foreign genes in prokaryotes. In *E. coli*, the ribosome binding site is called the Shine-Dalgarno (SD) sequence and includes an initiation codon (ATG) and a sequence 3-9 nucleotides in length located 3-11 nucleotides upstream of the initiation codon [Shine *et al.* (1975) *Nature* 254:34]. The SD sequence is thought to promote binding of mRNA to the ribosome by the pairing of bases between the SD sequence and the 3' end of *E. coli* 16S rRNA [Steitz *et al.* (1979) "Genetic signals and nucleotide sequences in messenger RNA." In *Biological Regulation and Development: Gene Expression* (ed. R.F. Goldberger)]. To express eukaryotic genes and prokaryotic genes with weak ribosome-binding site [Sambrook *et al.* (1989) "Expression of cloned genes in *Escherichia coli*." In *Molecular Cloning: A Laboratory Manual*].

A DNA molecule may be expressed intracellularly. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus will always be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide or by either *in vivo* or *in vitro* incubation with a bacterial methionine N-terminal peptidase (EPO-A-0 219 237).

Fusion proteins provide an alternative to direct expression. Usually, a DNA sequence encoding the N-terminal portion of an endogenous bacterial protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the bacteriophage lambda cell gene can be linked at the 5' terminus of a foreign gene and expressed in bacteria. The resulting fusion protein preferably retains a site for a processing enzyme (factor Xa) to cleave the bacteriophage protein from the foreign gene [Nagai *et al.* (1984) *Nature* 309:810]. Fusion proteins can also be made with sequences from the *lacZ* [Jia *et al.* (1987) *Gene* 60:197], *trpE* [Allen *et al.* (1987) *J. Biotechnol.* 5:93; Makoff *et al.* (1989) *J. Gen. Microbiol.* 135:11], and *Chey* [EP-A-0 324 647] genes. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (*eg.* ubiquitin specific processing-protease) to cleave the ubiquitin from the foreign protein. Through this method, native foreign protein can be isolated [Miller *et al.* (1989) *Bio/Technology* 7:698].

Alternatively, foreign proteins can also be secreted from the cell by creating chimeric DNA molecules that encode a fusion protein comprised of a signal peptide sequence fragment that provides for secretion of the foreign protein in bacteria [US patent 4,336,336]. The signal sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The protein is either secreted into the growth media (gram-positive bacteria) or into the periplasmic space, located between the inner and outer membrane of the cell (gram-negative bacteria). Preferably there are processing sites, which can be cleaved either *in vivo* or *in vitro* encoded between the signal peptide fragment and the foreign gene.

DNA encoding suitable signal sequences can be derived from genes for secreted bacterial proteins, such as the *E. coli* outer membrane protein gene (*ompA*) [Masui *et al.* (1983), in: *Experimental Manipulation of Gene Expression*; Ghrayeb *et al.* (1984) *EMBO J.* 3:2437] and the *E. coli* alkaline phosphatase signal sequence (*phoA*) [Oka *et al.* (1985) *Proc. Natl. Acad. Sci.* 82:7212]. As an additional example, the signal sequence of the alpha-amylase gene from various *Bacillus* strains can be used to secrete heterologous proteins from *B. subtilis* [Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 244 042].

Usually, transcription termination sequences recognized by bacteria are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Transcription termination sequences frequently include DNA sequences of about 50 nucleotides capable of forming stem loop structures that aid in terminating transcription. Examples include transcription termination sequences derived from genes with strong promoters, such as the *trp* gene in *E. coli* as well as other biosynthetic genes.

Usually, the above described components, comprising a promoter, signal sequence (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (eg. plasmids) capable of stable maintenance in a host, such as bacteria. The replicon will have a replication system, thus allowing it to be maintained in a prokaryotic host either for expression or for cloning and amplification. In addition, a replicon may be either a high or low copy number plasmid. A high copy number plasmid will generally have a copy number ranging from about 5 to about 200, and usually about 10 to about 150. A host containing a high copy

number plasmid will preferably contain at least about 10, and more preferably at least about 20 plasmids. Either a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host.

Alternatively, the expression constructs can be integrated into the bacterial genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to the bacterial chromosome that allows the vector to integrate. Integrations appear to result from recombinations between homologous DNA in the vector and the bacterial chromosome. For example, integrating vectors constructed with DNA from various *Bacillus* strains integrate into the *Bacillus* chromosome (EP-A- 0 127 328). Integrating vectors may also be comprised of bacteriophage or transposon sequences.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of bacterial strains that have been transformed. Selectable markers can be expressed in the bacterial host and may include genes which render bacteria resistant to drugs such as ampicillin, chloramphenicol, erythromycin, kanamycin (neomycin), and tetracycline [Davies *et al.* (1978) *Annu. Rev. Microbiol.* 32:469]. Selectable markers may also include biosynthetic genes, such as those in the histidine, tryptophan, and leucine biosynthetic pathways.

Alternatively, some of the above described components can be put together in transformation vectors. Transformation vectors are usually comprised of a selectable market that is either maintained in a replicon or developed into an integrating vector, as described above.

Expression and transformation vectors, either extra-chromosomal replicons or integrating vectors, have been developed for transformation into many bacteria. For example, expression vectors have been developed for, *inter alia*, the following bacteria: *Bacillus subtilis* [Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541], *Escherichia coli* [Shimatake *et al.* (1981) *Nature* 292:128; Amann *et al.* (1985) *Gene* 40:183; Studier *et al.* (1986) *J. Mol. Biol.* 189:113; EP-A-0 036 776, EP-A-0 136 829 and EP-A-0 136 907], *Streptococcus cremoris* [Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655]; *Streptococcus lividans* [Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655], *Streptomyces lividans* [US patent 4,745,056].

- Methods of introducing exogenous DNA into bacterial hosts are well-known in the art, and usually include either the transformation of bacteria treated with  $\text{CaCl}_2$  or other agents, such as divalent cations and DMSO. DNA can also be introduced into bacterial cells by electroporation. Transformation procedures usually vary with the bacterial species to be transformed. See *eg.*
- 5 [Masson *et al.* (1989) *FEMS Microbiol. Lett.* 60:273; Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541, *Bacillus*], [Miller *et al.* (1988) *Proc. Natl. Acad. Sci.* 85:856; Wang *et al.* (1990) *J. Bacteriol.* 172:949, *Campylobacter*], [Cohen *et al.* (1973) *Proc. Natl. Acad. Sci.* 69:2110; Dower *et al.* (1988) *Nucleic Acids Res.* 16:6127; Kushner (1978) "An improved method for transformation of *Escherichia coli* with ColE1-derived
- 10 plasmids. In *Genetic Engineering: Proceedings of the International Symposium on Genetic Engineering* (eds. H.W. Boyer and S. Nicosia); Mandel *et al.* (1970) *J. Mol. Biol.* 53:159; Taketo (1988) *Biochim. Biophys. Acta* 949:318; *Escherichia*], [Chassy *et al.* (1987) *FEMS Microbiol. Lett.* 44:173 *Lactobacillus*]; [Fiedler *et al.* (1988) *Anal. Biochem* 170:38, *Pseudomonas*]; [Augustin *et al.* (1990) *FEMS Microbiol. Lett.* 66:203, *Staphylococcus*], [Barany *et al.* (1980) *J. Bacteriol.*
- 15 144:698; Harlander (1987) "Transformation of *Streptococcus lactis* by electroporation, in: *Streptococcal Genetics* (ed. J. Ferretti and R. Curtiss III); Perry *et al.* (1981) *Infect. Immun.* 32:1295; Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655; Somkuti *et al.* (1987) *Proc. 4th Evr. Cong. Biotechnology* 1:412, *Streptococcus*].

#### v. Yeast Expression

- 20 Yeast expression systems are also known to one of ordinary skill in the art. A yeast promoter is any DNA sequence capable of binding yeast RNA polymerase and initiating the downstream (3') transcription of a coding sequence (*eg.* structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site (the "TATA
- 25 Box") and a transcription initiation site. A yeast promoter may also have a second domain called an upstream activator sequence (UAS), which, if present, is usually distal to the structural gene. The UAS permits regulated (inducible) expression. Constitutive expression occurs in the absence of a UAS. Regulated expression may be either positive or negative, thereby either enhancing or reducing transcription.
- 30 Yeast is a fermenting organism with an active metabolic pathway, therefore sequences encoding enzymes in the metabolic pathway provide particularly useful promoter sequences. Examples

include alcohol dehydrogenase (ADH) (EP-A-0 284 044), enolase, glucokinase, glucose-6-phosphate isomerase, glyceraldehyde-3-phosphate-dehydrogenase (GAP or GAPDH), hexokinase, phosphofructokinase, 3-phosphoglycerate mutase, and pyruvate kinase (PyK) (EPO-A-0 329 203). The yeast *PHO5* gene, encoding acid phosphatase, also provides useful promoter sequences

5 [Myanohara *et al.* (1983) *Proc. Natl. Acad. Sci. USA* 80:1].

In addition, synthetic promoters which do not occur in nature also function as yeast promoters. For example, UAS sequences of one yeast promoter may be joined with the transcription activation region of another yeast promoter, creating a synthetic hybrid promoter. Examples of such hybrid promoters include the ADH regulatory sequence linked to the GAP transcription activation region

10 (US Patent Nos. 4,876,197 and 4,880,734). Other examples of hybrid promoters include promoters which consist of the regulatory sequences of either the *ADH2*, *GAL4*, *GAL10*, OR *PHO5* genes, combined with the transcriptional activation region of a glycolytic enzyme gene such as GAP or PyK (EP-A-0 164 556). Furthermore, a yeast promoter can include naturally occurring promoters of non-yeast origin that have the ability to bind yeast RNA polymerase and initiate transcription.

15 Examples of such promoters include, *inter alia*, [Cohen *et al.* (1980) *Proc. Natl. Acad. Sci. USA* 77:1078; Henikoff *et al.* (1981) *Nature* 283:835; Hollenberg *et al.* (1981) *Curr. Topics Microbiol. Immunol.* 96:119; Hollenberg *et al.* (1979) "The Expression of Bacterial Antibiotic Resistance Genes in the Yeast *Saccharomyces cerevisiae*," in: *Plasmids of Medical, Environmental and Commercial Importance* (eds. K.N. Timmis and A. Puhler); Mercerau-Puigalon *et al.* (1980) *Gene*

20 11:163; Panthier *et al.* (1980) *Curr. Genet.* 2:109;].

A DNA molecule may be expressed intracellularly in yeast. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein will always be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with

25 cyanogen bromide.

Fusion proteins provide an alternative for yeast expression systems, as well as in mammalian, baculovirus, and bacterial expression systems. Usually, a DNA sequence encoding the N-terminal portion of an endogenous yeast protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two

30 amino acid sequences. For example, the yeast or human superoxide dismutase (SOD) gene, can be

linked at the 5' terminus of a foreign gene and expressed in yeast. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. See *eg.* EP-A-0 196 056. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (*eg.* ubiquitin-specific processing protease) to cleave the ubiquitin from the foreign protein. Through this method, therefore, native foreign protein can be isolated (*eg.* WO88/024066).

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provide for secretion in yeast of the foreign protein. Preferably, there are processing sites encoded between the leader fragment and the foreign gene that can be cleaved either *in vivo* or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell.

DNA encoding suitable signal sequences can be derived from genes for secreted yeast proteins, such as the yeast invertase gene (EP-A-0 012 873; JPO. 62,096,086) and the A-factor gene (US patent 4,588,684). Alternatively, leaders of non-yeast origin, such as an interferon leader, exist that also provide for secretion in yeast (EP-A-0 060 057).

A preferred class of secretion leaders are those that employ a fragment of the yeast alpha-factor gene, which contains both a "pre" signal sequence, and a "pro" region. The types of alpha-factor fragments that can be employed include the full-length pre-pro alpha factor leader (about 83 amino acid residues) as well as truncated alpha-factor leaders (usually about 25 to about 50 amino acid residues) (US Patents 4,546,083 and 4,870,008; EP-A-0 324 274). Additional leaders employing an alpha-factor leader fragment that provides for secretion include hybrid alpha-factor leaders made with a presequence of a first yeast, but a pro-region from a second yeast alphafactor. (*eg.* see WO 89/02463.)

Usually, transcription termination sequences recognized by yeast are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminator sequence and other yeast-recognized termination sequences, such as those coding for glycolytic enzymes.

Usually, the above described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (eg. plasmids) capable of stable maintenance in a host, such as yeast or bacteria. The replicon may have two replication systems, thus allowing it to be maintained, for example, in yeast for expression and in a prokaryotic host for cloning and amplification. Examples of such yeast-bacteria shuttle vectors include YEp24 [Botstein *et al.* (1979) *Gene* 8:17-24], pCI/1 [Brake *et al.* (1984) *Proc. Natl. Acad. Sci USA* 81:4642-4646], and YRp17 [Stinchcomb *et al.* (1982) *J. Mol. Biol.* 158:157]. In addition, a replicon may be either a high or low copy number plasmid. A high copy number plasmid will generally have a copy number ranging from about 5 to about 200, and usually about 10 to about 150. A host containing a high copy number plasmid will preferably have at least about 10, and more preferably at least about 20. Enter a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host. See eg. Brake *et al.*, *supra*.

Alternatively, the expression constructs can be integrated into the yeast genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to a yeast chromosome that allows the vector to integrate, and preferably contain two homologous sequences flanking the expression construct. Integrations appear to result from recombinations between homologous DNA in the vector and the yeast chromosome [Orr-Weaver *et al.* (1983) *Methods in Enzymol.* 101:228-245]. An integrating vector may be directed to a specific locus in yeast by selecting the appropriate homologous sequence for inclusion in the vector. See Orr-Weaver *et al.*, *supra*. One or more expression construct may integrate, possibly affecting levels of recombinant protein produced [Rine *et al.* (1983) *Proc. Natl. Acad. Sci. USA* 80:6750]. The chromosomal sequences included in the vector can occur either as a single segment in the vector, which results in the integration of the entire vector, or two segments homologous to adjacent segments in the chromosome and flanking the expression construct in the vector, which can result in the stable integration of only the expression construct.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of yeast strains that have been transformed. Selectable markers may include biosynthetic genes that can be expressed in the yeast host, such as *ADE2*, *HIS4*, *LEU2*,



*TRP1*, and *ALG7*, and the G418 resistance gene, which confer resistance in yeast cells to tunicamycin and G418, respectively. In addition, a suitable selectable marker may also provide yeast with the ability to grow in the presence of toxic compounds, such as metal. For example, the presence of *CUP1* allows yeast to grow in the presence of copper ions [Butt *et al.* (1987) *Microbiol. Rev.* 51:351].

Alternatively, some of the above described components can be put together into transformation vectors. Transformation vectors are usually comprised of a selectable marker that is either maintained in a replicon or developed into an integrating vector, as described above.

Expression and transformation vectors, either extrachromosomal replicons or integrating vectors, have been developed for transformation into many yeasts. For example, expression vectors have been developed for, *inter alia*, the following yeasts: *Candida albicans* [Kurtz, *et al.* (1986) *Mol. Cell. Biol.* 6:142], *Candida maltosa* [Kunze, *et al.* (1985) *J. Basic Microbiol.* 25:141], *Hansenula polymorpha* [Gleeson, *et al.* (1986) *J. Gen. Microbiol.* 132:3459; Roggenkamp *et al.* (1986) *Mol. Gen. Genet.* 202:302], *Kluyveromyces fragilis* [Das, *et al.* (1984) *J. Bacteriol.* 158:1165], *Kluyveromyces lactis* [De Louvencourt *et al.* (1983) *J. Bacteriol.* 154:737; Van den Berg *et al.* (1990) *Bio/Technology* 8:135], *Pichia guilliermondii* [Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141], *Pichia pastoris* [Cregg, *et al.* (1985) *Mol. Cell. Biol.* 5:3376; US Patent Nos. 4,837,148 and 4,929,555], *Saccharomyces cerevisiae* [Hinnen *et al.* (1978) *Proc. Natl. Acad. Sci. USA* 75:1929; Ito *et al.* (1983) *J. Bacteriol.* 153:163], *Schizosaccharomyces pombe* [Beach and Nurse (1981) *Nature* 300:706], and *Yarrowia lipolytica* [Davidow, *et al.* (1985) *Curr. Genet.* 10:380471 Gaillardin, *et al.* (1985) *Curr. Genet.* 10:49].

Methods of introducing exogenous DNA into yeast hosts are well-known in the art, and usually include either the transformation of spheroplasts or of intact yeast cells treated with alkali cations. Transformation procedures usually vary with the yeast species to be transformed. See *eg.* [Kurtz *et al.* (1986) *Mol. Cell. Biol.* 6:142; Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141; *Candida*]; [Gleeson *et al.* (1986) *J. Gen. Microbiol.* 132:3459; Roggenkamp *et al.* (1986) *Mol. Gen. Genet.* 202:302; *Hansenula*]; [Das *et al.* (1984) *J. Bacteriol.* 158:1165; De Louvencourt *et al.* (1983) *J. Bacteriol.* 154:1165; Van den Berg *et al.* (1990) *Bio/Technology* 8:135; *Kluyveromyces*]; [Cregg *et al.* (1985) *Mol. Cell. Biol.* 5:3376; Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141; US Patent Nos. 4,837,148 and 4,929,555; *Pichia*]; [Hinnen *et al.* (1978) *Proc. Natl. Acad. Sci. USA* 75:1929;

Ito *et al.* (1983) *J. Bacteriol.* 153:163 *Saccharomyces*]; [Beach and Nurse (1981) *Nature* 300:706; *Schizosaccharomyces*]; [Davidow *et al.* (1985) *Curr. Genet.* 10:39; Gaillardin *et al.* (1985) *Curr. Genet.* 10:49; *Yarrowia*].

### Antibodies

- 5 As used herein, the term "antibody" refers to a polypeptide or group of polypeptides composed of at least one antibody combining site. An "antibody combining site" is the three-dimensional binding space with an internal surface shape and charge distribution complementary to the features of an epitope of an antigen, which allows a binding of the antibody with the antigen. "Antibody" includes, for example, vertebrate antibodies, hybrid antibodies, chimeric antibodies, humanised  
10 antibodies, altered antibodies, univalent antibodies, Fab proteins, and single domain antibodies.

Antibodies against the proteins of the invention are useful for affinity chromatography, immunoassays, and distinguishing/identifying Neisserial proteins.

- Antibodies to the proteins of the invention, both polyclonal and monoclonal, may be prepared by conventional methods. In general, the protein is first used to immunize a suitable animal, preferably  
15 a mouse, rat, rabbit or goat. Rabbits and goats are preferred for the preparation of polyclonal sera due to the volume of serum obtainable, and the availability of labeled anti-rabbit and anti-goat antibodies. Immunization is generally performed by mixing or emulsifying the protein in saline, preferably in an adjuvant such as Freund's complete adjuvant, and injecting the mixture or emulsion parenterally (generally subcutaneously or intramuscularly). A dose of 50-200 µg/injection  
20 is typically sufficient. Immunization is generally boosted 2-6 weeks later with one or more injections of the protein in saline, preferably using Freund's incomplete adjuvant. One may alternatively generate antibodies by *in vitro* immunization using methods known in the art, which for the purposes of this invention is considered equivalent to *in vivo* immunization. Polyclonal antisera is obtained by bleeding the immunized animal into a glass or plastic container, incubating  
25 the blood at 25°C for one hour, followed by incubating at 4°C for 2-18 hours. The serum is recovered by centrifugation (eg. 1,000g for 10 minutes). About 20-50 ml per bleed may be obtained from rabbits.

Monoclonal antibodies are prepared using the standard method of Kohler & Milstein [*Nature* (1975) 256:495-96], or a modification thereof. Typically, a mouse or rat is immunized as described

above. However, rather than bleeding the animal to extract serum, the spleen (and optionally several large lymph nodes) is removed and dissociated into single cells. If desired, the spleen cells may be screened (after removal of nonspecifically adherent cells) by applying a cell suspension to a plate or well coated with the protein antigen. B-cells expressing membrane-bound immunoglobulin specific for the antigen bind to the plate, and are not rinsed away with the rest of the suspension. Resulting B-cells, or all dissociated spleen cells, are then induced to fuse with myeloma cells to form hybridomas, and are cultured in a selective medium (eg. hypoxanthine, aminopterin, thymidine medium, "HAT"). The resulting hybridomas are plated by limiting dilution, and are assayed for the production of antibodies which bind specifically to the immunizing antigen (and which do not bind to unrelated antigens). The selected MAb-secreting hybridomas are then cultured either *in vitro* (eg. in tissue culture bottles or hollow fiber reactors), or *in vivo* (as ascites in mice).

If desired, the antibodies (whether polyclonal or monoclonal) may be labeled using conventional techniques. Suitable labels include fluorophores, chromophores, radioactive atoms (particularly  $^{32}\text{P}$  and  $^{125}\text{I}$ ), electron-dense reagents, enzymes, and ligands having specific binding partners. Enzymes are typically detected by their activity. For example, horseradish peroxidase is usually detected by its ability to convert 3,3',5,5'-tetramethylbenzidine (TMB) to a blue pigment, quantifiable with a spectrophotometer. "Specific binding partner" refers to a protein capable of binding a ligand molecule with high specificity, as for example in the case of an antigen and a monoclonal antibody specific therefor. Other specific binding partners include biotin and avidin or streptavidin, IgG and protein A, and the numerous receptor-ligand couples known in the art. It should be understood that the above description is not meant to categorize the various labels into distinct classes, as the same label may serve in several different modes. For example,  $^{125}\text{I}$  may serve as a radioactive label or as an electron-dense reagent. HRP may serve as enzyme or as antigen for a MAb. Further, one may combine various labels for desired effect. For example, MAbs and avidin also require labels in the practice of this invention: thus, one might label a MAb with biotin, and detect its presence with avidin labeled with  $^{125}\text{I}$ , or with an anti-biotin MAb labeled with HRP. Other permutations and possibilities will be readily apparent to those of ordinary skill in the art, and are considered as equivalents within the scope of the instant invention.

Pharmaceutical Compositions

Pharmaceutical compositions can comprise either polypeptides, antibodies, or nucleic acid of the invention. The pharmaceutical compositions will comprise a therapeutically effective amount of either polypeptides, antibodies, or polynucleotides of the claimed invention.

- 5 The term "therapeutically effective amount" as used herein refers to an amount of a therapeutic agent to treat, ameliorate, or prevent a desired disease or condition, or to exhibit a detectable therapeutic or preventative effect. The effect can be detected by, for example, chemical markers or antigen levels. Therapeutic effects also include reduction in physical symptoms, such as decreased body temperature. The precise effective amount for a subject will depend upon the subject's size  
10 and health, the nature and extent of the condition, and the therapeutics or combination of therapeutics selected for administration. Thus, it is not useful to specify an exact effective amount in advance. However, the effective amount for a given situation can be determined by routine experimentation and is within the judgement of the clinician.

- For purposes of the present invention, an effective dose will be from about 0.01 mg/kg to 50 mg/kg  
15 or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

- A pharmaceutical composition can also contain a pharmaceutically acceptable carrier. The term "pharmaceutically acceptable carrier" refers to a carrier for administration of a therapeutic agent, such as antibodies or a polypeptide, genes, and other therapeutic agents. The term refers to any  
20 pharmaceutical carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition, and which may be administered without undue toxicity. Suitable carriers may be large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, and inactive virus particles. Such carriers are well known to those of ordinary skill in  
25 the art.

Pharmaceutically acceptable salts can be used therein, for example, mineral acid salts such as hydrochlorides, hydrobromides, phosphates, sulfates, and the like; and the salts of organic acids such as acetates, propionates, malonates, benzoates, and the like. A thorough discussion of

pharmaceutically acceptable excipients is available in Remington's Pharmaceutical Sciences (Mack Pub. Co., N.J. 1991).

Pharmaceutically acceptable carriers in therapeutic compositions may contain liquids such as water, saline, glycerol and ethanol. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles. Typically, the therapeutic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. Liposomes are included within the definition of a pharmaceutically acceptable carrier.

#### Delivery Methods

Once formulated, the compositions of the invention can be administered directly to the subject. The subjects to be treated can be animals; in particular, human subjects can be treated.

Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal or transcutaneous applications (eg. see WO98/20734), needles, and gene guns or hyposprays. Dosage treatment may be a single dose schedule or a multiple dose schedule.

#### Vaccines

Vaccines according to the invention may either be prophylactic (*ie.* to prevent infection) or therapeutic (*ie.* to treat disease after infection).

Such vaccines comprise immunising antigen(s), immunogen(s), polypeptide(s), protein(s) or nucleic acid, usually in combination with "pharmaceutically acceptable carriers," which include any carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, lipid aggregates (such as oil droplets or liposomes), and inactive virus particles. Such carriers are well known to those of ordinary skill in the art. Additionally, these carriers may function as immunostimulating agents ("adjuvants"). Furthermore, the antigen or immunogen may be conjugated to a bacterial toxoid, such as a toxoid from diphtheria, tetanus, cholera, *H. pylori*, etc. pathogens.

Preferred adjuvants to enhance effectiveness of the composition include, but are not limited to: (1) aluminum salts (alum), such as aluminum hydroxide, aluminum phosphate, aluminum sulfate, etc; (2) oil-in-water emulsion formulations (with or without other specific immunostimulating agents such as muramyl peptides (see below) or bacterial cell wall components), such as for example (a) MF59™ (WO 90/14837; Chapter 10 in *Vaccine design: the subunit and adjuvant approach*, eds. Powell & Newman, Plenum Press 1995), containing 5% Squalene, 0.5% Tween 80, and 0.5% Span 85 (optionally containing various amounts of MTP-PE (see below), although not required) formulated into submicron particles using a microfluidizer such as Model 110Y microfluidizer (Microfluidics, Newton, MA), (b) SAF, containing 10% Squalane, 0.4% Tween 80, 5% pluronic-blocked polymer L121, and thr-MDP (see below) either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion, and (c) Ribi™ adjuvant system (RAS), (Ribi Immunochem, Hamilton, MT) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial cell wall components from the group consisting of monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (Detox™); (3) saponin adjuvants, such as Stimulon™ (Cambridge Bioscience, Worcester, MA) may be used or particles generated therefrom such as ISCOMs (immunostimulating complexes); (4) Complete Freund's Adjuvant (CFA) and Incomplete Freund's Adjuvant (IFA); (5) cytokines, such as interleukins (eg. IL-1, IL-2, IL-4, IL-5, IL-6, IL-7, IL-12, etc.), interferons (eg. gamma interferon), macrophage colony stimulating factor (M-CSF), tumor necrosis factor (TNF), etc; and (6) other substances that act as immunostimulating agents to enhance the effectiveness of the composition. Alum and MF59™ are preferred.

As mentioned above, muramyl peptides include, but are not limited to, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-normuramyl-L-alanyl-D-isoglutamine (nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-*sn*-glycero-3-hydroxyphosphoryloxy)-ethylamine (MTP-PE), etc.

The immunogenic compositions (eg. the immunising antigen/immunogen/polypeptide/protein/nucleic acid, pharmaceutically acceptable carrier, and adjuvant) typically will contain diluents, such as water, saline, glycerol, ethanol, etc. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles.

Typically, the immunogenic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. The preparation also may be emulsified or encapsulated in liposomes for enhanced adjuvant effect, as discussed above under pharmaceutically acceptable carriers.

- 5 Immunogenic compositions used as vaccines comprise an immunologically effective amount of the antigenic or immunogenic polypeptides, as well as any other of the above-mentioned components, as needed. By "immunologically effective amount", it is meant that the administration of that amount to an individual, either in a single dose or as part of a series, is effective for treatment or prevention. This amount varies depending upon the health and physical condition of the individual
- 10 to be treated, the taxonomic group of individual to be treated (*eg.* nonhuman primate, primate, *etc.*), the capacity of the individual's immune system to synthesize antibodies, the degree of protection desired, the formulation of the vaccine, the treating doctor's assessment of the medical situation, and other relevant factors. It is expected that the amount will fall in a relatively broad range that can be determined through routine trials.
- 15 The immunogenic compositions are conventionally administered parenterally, *eg.* by injection, either subcutaneously, intramuscularly, or transdermally/transcutaneously (*eg.* WO98/20734). Additional formulations suitable for other modes of administration include oral and pulmonary formulations, suppositories, and transdermal applications. Dosage treatment may be a single dose schedule or a multiple dose schedule. The vaccine may be administered in conjunction with other
- 20 immunoregulatory agents.

As an alternative to protein-based vaccines, DNA vaccination may be employed [*eg.* Robinson & Torres (1997) *Seminars in Immunology* 9:271-283; Donnelly *et al.* (1997) *Annu Rev Immunol* 15:617-648; see later herein].

#### Gene Delivery Vehicles

- 25 Gene therapy vehicles for delivery of constructs including a coding sequence of a therapeutic of the invention, to be delivered to the mammal for expression in the mammal, can be administered either locally or systemically. These constructs can utilize viral or non-viral vector approaches in *in vivo* or *ex vivo* modality. Expression of such coding sequence can be induced using endogenous

mammalian or heterologous promoters. Expression of the coding sequence in vivo can be either constitutive or regulated.

The invention includes gene delivery vehicles capable of expressing the contemplated nucleic acid sequences. The gene delivery vehicle is preferably a viral vector and, more preferably, a retroviral, adenoviral, adeno-associated viral (AAV), herpes viral, or alphavirus vector. The viral vector can also be an astrovirus, coronavirus, orthomyxovirus, papovavirus, paramyxovirus, parvovirus, picornavirus, poxvirus, or togavirus viral vector. See generally, Jolly (1994) *Cancer Gene Therapy* 1:51-64; Kimura (1994) *Human Gene Therapy* 5:845-852; Connelly (1995) *Human Gene Therapy* 6:185-193; and Kaplitt (1994) *Nature Genetics* 6:148-153.

10 Retroviral vectors are well known in the art and we contemplate that any retroviral gene therapy vector is employable in the invention, including B, C and D type retroviruses, xenotropic retroviruses (for example, NZB-X1, NZB-X2 and NZB9-1 (see O'Neill (1985) *J. Virol.* 53:160) polytropic retroviruses eg. MCF and MCF-MLV (see Kelly (1983) *J. Virol.* 45:291), spumaviruses and lentiviruses. See RNA Tumor Viruses, Second Edition, Cold Spring Harbor Laboratory, 1985.

15 Portions of the retroviral gene therapy vector may be derived from different retroviruses. For example, retrovector LTRs may be derived from a Murine Sarcoma Virus, a tRNA binding site from a Rous Sarcoma Virus, a packaging signal from a Murine Leukemia Virus, and an origin of second strand synthesis from an Avian Leukosis Virus.

20 These recombinant retroviral vectors may be used to generate transduction competent retroviral vector particles by introducing them into appropriate packaging cell lines (see US patent 5,591,624). Retrovirus vectors can be constructed for site-specific integration into host cell DNA by incorporation of a chimeric integrase enzyme into the retroviral particle (see WO96/37626). It is preferable that the recombinant viral vector is a replication defective recombinant virus.

25 Packaging cell lines suitable for use with the above-described retrovirus vectors are well known in the art, are readily prepared (see WO95/30763 and WO92/05266), and can be used to create producer cell lines (also termed vector cell lines or "VCLs") for the production of recombinant vector particles. Preferably, the packaging cell lines are made from human parent cells (eg. HT1080 cells) or mink parent cell lines, which eliminates inactivation in human serum.



Preferred retroviruses for the construction of retroviral gene therapy vectors include Avian Leukosis Virus, Bovine Leukemia Virus, Murine Leukemia Virus, Mink-Cell Focus-Inducing Virus, Murine Sarcoma Virus, Reticuloendotheliosis Virus and Rous Sarcoma Virus. Particularly preferred Murine Leukemia Viruses include 4070A and 1504A (Hartley and Rowe (1976) *J Virol* 19:19-25), Abelson (ATCC No. VR-999), Friend (ATCC No. VR-245), Graffi, Gross (ATCC No. VR-590), Kirsten, Harvey Sarcoma Virus and Rauscher (ATCC No. VR-998) and Moloney Murine Leukemia Virus (ATCC No. VR-190). Such retroviruses may be obtained from depositories or collections such as the American Type Culture Collection ("ATCC") in Rockville, Maryland or isolated from known sources using commonly available techniques.

- Exemplary known retroviral gene therapy vectors employable in this invention include those described in patent applications GB2200651, EP0415731, EP0345242, EP0334301, WO89/02468; WO89/05349, WO89/09271, WO90/02806, WO90/07936, WO94/03622, WO93/25698, WO93/25234, WO93/11230, WO93/10218, WO91/02805, WO91/02825, WO95/07994, US 5,219,740, US 4,405,712, US 4,861,719, US 4,980,289, US 4,777,127, US 5,591,624. See also Vile (1993) *Cancer Res* 53:3860-3864; Vile (1993) *Cancer Res* 53:962-967; Ram (1993) *Cancer Res* 53 (1993) 83-88; Takamiya (1992) *J Neurosci Res* 33:493-503; Baba (1993) *J Neurosurg* 79:729-735; Mann (1983) *Cell* 33:153; Cane (1984) *Proc Natl Acad Sci* 81:6349; and Miller (1990) *Human Gene Therapy* 1.

- Human adenoviral gene therapy vectors are also known in the art and employable in this invention. See, for example, Berkner (1988) *Biotechniques* 6:616 and Rosenfeld (1991) *Science* 252:431, and WO93/07283, WO93/06223, and WO93/07282. Exemplary known adenoviral gene therapy vectors employable in this invention include those described in the above referenced documents and in WO94/12649, WO93/03769, WO93/19191, WO94/28938, WO95/11984, WO95/00655, WO95/27071, WO95/29993, WO95/34671, WO96/05320, WO94/08026, WO94/11506, WO93/06223, WO94/24299, WO95/14102, WO95/24297, WO95/02697, WO94/28152, WO94/24299, WO95/09241, WO95/25807, WO95/05835, WO94/18922 and WO95/09654. Alternatively, administration of DNA linked to killed adenovirus as described in Curiel (1992) *Hum. Gene Ther.* 3:147-154 may be employed. The gene delivery vehicles of the invention also include adenovirus associated virus (AAV) vectors. Leading and preferred examples of such vectors for use in this invention are the AAV-2 based vectors disclosed in Srivastava,

WO93/09239. Most preferred AAV vectors comprise the two AAV inverted terminal repeats in which the native D-sequences are modified by substitution of nucleotides, such that at least 5 native nucleotides and up to 18 native nucleotides, preferably at least 10 native nucleotides up to 18 native nucleotides, most preferably 10 native nucleotides are retained and the remaining nucleotides of the D-sequence are deleted or replaced with non-native nucleotides. The native D-sequences of the AAV inverted terminal repeats are sequences of 20 consecutive nucleotides in each AAV inverted terminal repeat (*ie.* there is one sequence at each end) which are not involved in HP formation. The non-native replacement nucleotide may be any nucleotide other than the nucleotide found in the native D-sequence in the same position. Other employable exemplary AAV vectors are pWP-19, pWN-1, both of which are disclosed in Nahreini (1993) *Gene* 124:257-262. Another example of such an AAV vector is psub201 (see Samulski (1987) *J. Virol.* 61:3096). Another exemplary AAV vector is the Double-D ITR vector. Construction of the Double-D ITR vector is disclosed in US Patent 5,478,745. Still other vectors are those disclosed in Carter US Patent 4,797,368 and Muzyczka US Patent 5,139,941, Chartejee US Patent 5,474,935, and Kotin WO94/288157. Yet a further example of an AAV vector employable in this invention is SSV9AFABTKneo, which contains the AFP enhancer and albumin promoter and directs expression predominantly in the liver. Its structure and construction are disclosed in Su (1996) *Human Gene Therapy* 7:463-470. Additional AAV gene therapy vectors are described in US 5,354,678, US 5,173,414, US 5,139,941, and US 5,252,479.

The gene therapy vectors of the invention also include herpes vectors. Leading and preferred examples are herpes simplex virus vectors containing a sequence encoding a thymidine kinase polypeptide such as those disclosed in US 5,288,641 and EP0176170 (Roizman). Additional exemplary herpes simplex virus vectors include HFEM/ICP6-LacZ disclosed in WO95/04139 (Wistar Institute), pHSVlac described in Geller (1988) *Science* 241:1667-1669 and in WO90/09441 and WO92/07945, HSV Us3::pgC-lacZ described in Fink (1992) *Human Gene Therapy* 3:11-19 and HSV 7134, 2 RH 105 and GAL4 described in EP 0453242 (Breakefield), and those deposited with the ATCC as accession numbers ATCC VR-977 and ATCC VR-260.

Also contemplated are alpha virus gene therapy vectors that can be employed in this invention. Preferred alpha virus vectors are Sindbis viruses vectors. Togaviruses, Semliki Forest virus (ATCC VR-67; ATCC VR-1247), Middleberg virus (ATCC VR-370), Ross River virus (ATCC VR-373;

ATCC VR-1246), Venezuelan equine encephalitis virus (ATCC VR923; ATCC VR-1250; ATCC VR-1249; ATCC VR-532), and those described in US patents 5,091,309, 5,217,879, and WO92/10578. More particularly, those alpha virus vectors described in US Serial No. 08/405,627, filed March 15, 1995, WO94/21792, WO92/10578, WO95/07994, US 5,091,309 and US 5,217,879 are employable. Such alpha viruses may be obtained from depositories or collections such as the ATCC in Rockville, Maryland or isolated from known sources using commonly available techniques. Preferably, alphavirus vectors with reduced cytotoxicity are used (see USSN 08/679640).

DNA vector systems such as eukaryotic layered expression systems are also useful for expressing the nucleic acids of the invention. See WO95/07994 for a detailed description of eukaryotic layered expression systems. Preferably, the eukaryotic layered expression systems of the invention are derived from alphavirus vectors and most preferably from Sindbis viral vectors.

Other viral vectors suitable for use in the present invention include those derived from poliovirus, for example ATCC VR-58 and those described in Evans, *Nature* 339 (1989) 385 and Sabin (1973) *J. Biol. Standardization* 1:115; rhinovirus, for example ATCC VR-1110 and those described in Arnold (1990) *J Cell Biochem* L401; pox viruses such as canary pox virus or vaccinia virus, for example ATCC VR-111 and ATCC VR-2010 and those described in Fisher-Hoch (1989) *Proc Natl Acad Sci* 86:317; Flexner (1989) *Ann NY Acad Sci* 569:86, Flexner (1990) *Vaccine* 8:17; in US 4,603,112 and US 4,769,330 and WO89/01973; SV40 virus, for example ATCC VR-305 and those described in Mulligan (1979) *Nature* 277:108 and Madzak (1992) *J Gen Virol* 73:1533; influenza virus, for example ATCC VR-797 and recombinant influenza viruses made employing reverse genetics techniques as described in US 5,166,057 and in Enami (1990) *Proc Natl Acad Sci* 87:3802-3805; Enami & Palese (1991) *J Virol* 65:2711-2713 and Luytjes (1989) *Cell* 59:110, (see also McMichael (1983) *NEJ Med* 309:13, and Yap (1978) *Nature* 273:238 and *Nature* (1979) 277:108); human immunodeficiency virus as described in EP-0386882 and in Buchschacher (1992) *J. Virol.* 66:2731; measles virus, for example ATCC VR-67 and VR-1247 and those described in EP-0440219; Aura virus, for example ATCC VR-368; Bebaru virus, for example ATCC VR-600 and ATCC VR-1240; Cabassou virus, for example ATCC VR-922; Chikungunya virus, for example ATCC VR-64 and ATCC VR-1241; Fort Morgan Virus, for example ATCC VR-924; Getah virus, for example ATCC VR-369 and ATCC VR-1243; Kyzylagach virus, for example

ATCC VR-927; Mayaro virus, for example ATCC VR-66; Mucambo virus, for example ATCC VR-580 and ATCC VR-1244; Ndumu virus, for example ATCC VR-371; Pixuna virus, for example ATCC VR-372 and ATCC VR-1245; Tonate virus, for example ATCC VR-925; Trinitivirus, for example ATCC VR-469; Una virus, for example ATCC VR-374; Whataroa virus, for example ATCC VR-926; Y-62-33 virus, for example ATCC VR-375; O'Nyong virus, Eastern encephalitis virus, for example ATCC VR-65 and ATCC VR-1242; Western encephalitis virus, for example ATCC VR-70, ATCC VR-1251, ATCC VR-622 and ATCC VR-1252; and coronavirus, for example ATCC VR-740 and those described in Hamre (1966) *Proc Soc Exp Biol Med* 121:190.

Delivery of the compositions of this invention into cells is not limited to the above mentioned viral vectors. Other delivery methods and media may be employed such as, for example, nucleic acid expression vectors, polycationic condensed DNA linked or unlinked to killed adenovirus alone, for example see US Serial No. 08/366,787, filed December 30, 1994 and Curiel (1992) *Hum Gene Ther* 3:147-154 ligand linked DNA, for example see Wu (1989) *J Biol Chem* 264:16985-16987, eucaryotic cell delivery vehicles cells, for example see US Serial No.08/240,030, filed May 9, 1994, and US Serial No. 08/404,796, deposition of photopolymerized hydrogel materials, hand-held gene transfer particle gun, as described in US Patent 5,149,655, ionizing radiation as described in US5,206,152 and in WO92/11033, nucleic charge neutralization or fusion with cell membranes. Additional approaches are described in Philip (1994) *Mol Cell Biol* 14:2411-2418 and in Woffendin (1994) *Proc Natl Acad Sci* 91:1581-1585.

Particle mediated gene transfer may be employed, for example see US Serial No. 60/023,867. Briefly, the sequence can be inserted into conventional vectors that contain conventional control sequences for high level expression, and then incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations like polylysine, protamine, and albumin, linked to cell targeting ligands such as asialoorosomucoid, as described in Wu & Wu (1987) *J. Biol. Chem.* 262:4429-4432, insulin as described in Hucked (1990) *Biochem Pharmacol* 40:253-263, galactose as described in Plank (1992) *Bioconjugate Chem* 3:533-539, lactose or transferrin.

Naked DNA may also be employed. Exemplary naked DNA introduction methods are described in WO 90/11092 and US 5,580,859. Uptake efficiency may be improved using biodegradable latex beads. DNA coated latex beads are efficiently transported into cells after endocytosis initiation by the beads. The method may be improved further by treatment of the beads to increase

hydrophobicity and thereby facilitate disruption of the endosome and release of the DNA into the cytoplasm.

Liposomes that can act as gene delivery vehicles are described in US 5,422,120, WO95/13796, WO94/23697, WO91/14445 and EP-524,968. As described in USSN. 60/023,867, on non-viral delivery, the nucleic acid sequences encoding a polypeptide can be inserted into conventional vectors that contain conventional control sequences for high level expression, and then be incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations like polylysine, protamine, and albumin, linked to cell targeting ligands such as asialoorosomucoid, insulin, galactose, lactose, or transferrin. Other delivery systems include the use of liposomes to encapsulate DNA comprising the gene under the control of a variety of tissue-specific or ubiquitously-active promoters. Further non-viral delivery suitable for use includes mechanical delivery systems such as the approach described in Woffendin *et al* (1994) *Proc. Natl. Acad. Sci. USA* 91(24):11581-11585. Moreover, the coding sequence and the product of expression of such can be delivered through deposition of photopolymerized hydrogel materials. Other conventional methods for gene delivery that can be used for delivery of the coding sequence include, for example, use of hand-held gene transfer particle gun, as described in US 5,149,655; use of ionizing radiation for activating transferred gene, as described in US 5,206,152 and WO92/11033

Exemplary liposome and polycationic gene delivery vehicles are those described in US 5,422,120 and 4,762,915; in WO 95/13796; WO94/23697; and WO91/14445; in EP-0524968; and in Stryer, Biochemistry, pages 236-240 (1975) W.H. Freeman, San Francisco; Szoka (1980) *Biochem Biophys Acta* 600:1; Bayer (1979) *Biochem Biophys Acta* 550:464; Rivnay (1987) *Meth Enzymol* 149:119; Wang (1987) *Proc Natl Acad Sci* 84:7851; Plant (1989) *Anal Biochem* 176:420.

A polynucleotide composition can comprises therapeutically effective amount of a gene therapy vehicle, as the term is defined above. For purposes of the present invention, an effective dose will be from about 0.01 mg/ kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

#### Delivery Methods

Once formulated, the polynucleotide compositions of the invention can be administered (1) directly to the subject; (2) delivered *ex vivo*, to cells derived from the subject; or (3) *in vitro* for expression

of recombinant proteins. The subjects to be treated can be mammals or birds. Also, human subjects can be treated.

Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal or transcutaneous applications (eg. see WO98/20734), needles, and gene guns or hyposprays. Dosage treatment may be a single dose schedule or a multiple dose schedule.

Methods for the *ex vivo* delivery and reimplantation of transformed cells into a subject are known in the art and described in eg. WO93/14778. Examples of cells useful in *ex vivo* applications include, for example, stem cells, particularly hematopoietic, lymph cells, macrophages, dendritic cells, or tumor cells.

Generally, delivery of nucleic acids for both *ex vivo* and *in vitro* applications can be accomplished by the following procedures, for example, dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei, all well known in the art.

#### Polynucleotide and polypeptide pharmaceutical compositions

In addition to the pharmaceutically acceptable carriers and salts described above, the following additional agents can be used with polynucleotide and/or polypeptide compositions.

#### A. Polypeptides

One example are polypeptides which include, without limitation: asialoglycoprotein (ASOR); transferrin; asialoglycoproteins; antibodies; antibody fragments; ferritin; interleukins; interferons, granulocyte, macrophage colony stimulating factor (GM-CSF), granulocyte colony stimulating factor (G-CSF), macrophage colony stimulating factor (M-CSF), stem cell factor and erythropoietin. Viral antigens, such as envelope proteins, can also be used. Also, proteins from other invasive organisms, such as the 17 amino acid peptide from the circumsporozoite protein of plasmodium falciparum known as RII.

B.Hormones, Vitamins, etc.

Other groups that can be included are, for example: hormones, steroids, androgens, estrogens, thyroid hormone, or vitamins, folic acid.

C.Polyalkylenes, Polysaccharides, etc.

- 5 Also, polyalkylene glycol can be included with the desired polynucleotides/polypeptides. In a preferred embodiment, the polyalkylene glycol is polyethylene glycol. In addition, mono-, di-, or polysaccharides can be included. In a preferred embodiment of this aspect, the polysaccharide is dextran or DEAE-dextran. Also, chitosan and poly(lactide-co-glycolide)

D.Lipids, and Liposomes

- 10 The desired polynucleotide/polypeptide can also be encapsulated in lipids or packaged in liposomes prior to delivery to the subject or to cells derived therefrom.

Lipid encapsulation is generally accomplished using liposomes which are able to stably bind or entrap and retain nucleic acid. The ratio of condensed polynucleotide to lipid preparation can vary but will generally be around 1:1 (mg DNA:micromoles lipid), or more of lipid. For a review of the  
15 use of liposomes as carriers for delivery of nucleic acids, see, Hug and Sleight (1991) *Biochim. Biophys. Acta.* 1097:1-17; Straubinger (1983) *Meth. Enzymol.* 101:512-527.

- Liposomal preparations for use in the present invention include cationic (positively charged), anionic (negatively charged) and neutral preparations. Cationic liposomes have been shown to mediate intracellular delivery of plasmid DNA (Felgner (1987) *Proc. Natl. Acad. Sci. USA*  
20 84:7413-7416); mRNA (Malone (1989) *Proc. Natl. Acad. Sci. USA* 86:6077-6081); and purified transcription factors (Debs (1990) *J. Biol. Chem.* 265:10189-10192), in functional form.

- Cationic liposomes are readily available. For example, N[1-2,3-dioleoyloxy)propyl]-N,N,N-triethylammonium (DOTMA) liposomes are available under the trademark Lipofectin, from GIBCO BRL, Grand Island, NY. (See, also, Felgner *supra*). Other commercially available liposomes include  
25 transfectace (DDAB/DOPE) and DOTAP/DOPE (Boehringer). Other cationic liposomes can be prepared from readily available materials using techniques well known in the art. See, eg. Szoka (1978) *Proc. Natl. Acad. Sci. USA* 75:4194-4198; WO90/11092 for a description of the synthesis of DOTAP (1,2-bis(oleoyloxy)-3-(trimethylammonio)propane) liposomes.

Similarly, anionic and neutral liposomes are readily available, such as from Avanti Polar Lipids (Birmingham, AL), or can be easily prepared using readily available materials. Such materials include phosphatidyl choline, cholesterol, phosphatidyl ethanolamine, dioleoylphosphatidyl choline (DOPC), dioleoylphosphatidyl glycerol (DOPG), dioleoylphosphatidyl ethanolamine (DOPE), among others. These materials can also be mixed with the DOTMA and DOTAP starting materials in appropriate ratios. Methods for making liposomes using these materials are well known in the art.

The liposomes can comprise multilammellar vesicles (MLVs), small unilamellar vesicles (SUVs), or large unilamellar vesicles (LUVs). The various liposome-nucleic acid complexes are prepared using methods known in the art. See eg. Straubinger (1983) *Meth. Immunol.* 101:512-527; Szoka (1978) *Proc. Natl. Acad. Sci. USA* 75:4194-4198; Papahadjopoulos (1975) *Biochim. Biophys. Acta* 394:483; Wilson (1979) *Cell* 17:77; Deamer & Bangham (1976) *Biochim. Biophys. Acta* 443:629; Ostro (1977) *Biochem. Biophys. Res. Commun.* 76:836; Fraley (1979) *Proc. Natl. Acad. Sci. USA* 76:3348; Enoch & Strittmatter (1979) *Proc. Natl. Acad. Sci. USA* 76:145; Fraley (1980) *J. Biol. Chem.* (1980) 255:10431; Szoka & Papahadjopoulos (1978) *Proc. Natl. Acad. Sci. USA* 75:145; and Schaefer-Ridder (1982) *Science* 215:166.

#### E.Lipoproteins

In addition, lipoproteins can be included with the polynucleotide/polypeptide to be delivered. Examples of lipoproteins to be utilized include: chylomicrons, HDL, IDL, LDL, and VLDL. Mutants, fragments, or fusions of these proteins can also be used. Also, modifications of naturally occurring lipoproteins can be used, such as acetylated LDL. These lipoproteins can target the delivery of polynucleotides to cells expressing lipoprotein receptors. Preferably, if lipoproteins are including with the polynucleotide to be delivered, no other targeting ligand is included in the composition.

Naturally occurring lipoproteins comprise a lipid and a protein portion. The protein portion are known as apoproteins. At the present, apoproteins A, B, C, D, and E have been isolated and identified. At least two of these contain several proteins, designated by Roman numerals, AI, AII, AIV; CI, CII, CIII.

A lipoprotein can comprise more than one apoprotein. For example, naturally occurring chylomicrons comprises of A, B, C, and E, over time these lipoproteins lose A and acquire C and



E apoproteins. VLDL comprises A, B, C, and E apoproteins, LDL comprises apoprotein B; and HDL comprises apoproteins A, C, and E.

The amino acid of these apoproteins are known and are described in, for example, Breslow (1985) *Annu Rev. Biochem* 54:699; Law (1986) *Adv. Exp Med. Biol.* 151:162; Chen (1986) *J Biol Chem* 261:12918; Kane (1980) *Proc Natl Acad Sci USA* 77:2465; and Utermann (1984) *Hum Genet* 65:232.

Lipoproteins contain a variety of lipids including, triglycerides, cholesterol (free and esters), and phospholipids. The composition of the lipids varies in naturally occurring lipoproteins. For example, chylomicrons comprise mainly triglycerides. A more detailed description of the lipid content of naturally occurring lipoproteins can be found, for example, in *Meth. Enzymol.* 128 (1986). The composition of the lipids are chosen to aid in conformation of the apoprotein for receptor binding activity. The composition of lipids can also be chosen to facilitate hydrophobic interaction and association with the polynucleotide binding molecule.

Naturally occurring lipoproteins can be isolated from serum by ultracentrifugation, for instance. Such methods are described in *Meth. Enzymol. (supra)*; Pitas (1980) *J. Biochem.* 255:5454-5460 and Mahey (1979) *J Clin. Invest* 64:743-750. Lipoproteins can also be produced by *in vitro* or recombinant methods by expression of the apoprotein genes in a desired host cell. See, for example, Atkinson (1986) *Annu Rev Biophys Chem* 15:403 and Radding (1958) *Biochim Biophys Acta* 30:443. Lipoproteins can also be purchased from commercial suppliers, such as Biomedical Technologies, Inc., Stoughton, Massachusetts, USA. Further description of lipoproteins can be found in Zuckermann *et al.* PCT/US97/14465.

#### F. Polycationic Agents

Polycationic agents can be included, with or without lipoprotein, in a composition with the desired polynucleotide/polypeptide to be delivered.

Polycationic agents, typically, exhibit a net positive charge at physiological relevant pH and are capable of neutralizing the electrical charge of nucleic acids to facilitate delivery to a desired location. These agents have both *in vitro*, *ex vivo*, and *in vivo* applications. Polycationic agents can be used to deliver nucleic acids to a living subject either intramuscularly, subcutaneously, etc.

The following are examples of useful polypeptides as polycationic agents: polylysine, polyarginine, polyornithine, and protamine. Other examples include histones, protamines, human serum albumin, DNA binding proteins, non-histone chromosomal proteins, coat proteins from DNA viruses, such as (X174, transcriptional factors also contain domains that bind DNA and therefore may be useful as nucleic acid condensing agents. Briefly, transcriptional factors such as C/CEBP, c-jun, c-fos, AP-1, AP-2, AP-3, CPF, Prot-1, Sp-1, Oct-1, Oct-2, CREP, and TFIID contain basic domains that bind DNA sequences.

Organic polycationic agents include: spermine, spermidine, and putrescine.

The dimensions and of the physical properties of a polycationic agent can be extrapolated from the list above, to construct other polypeptide polycationic agents or to produce synthetic polycationic agents.

Synthetic polycationic agents which are useful include, for example, DEAE-dextran, polybrene. Lipofectin™, and lipofectAMINE™ are monomers that form polycationic complexes when combined with polynucleotides/polypeptides.

#### 15 Immunodiagnostic Assays

Neisserial antigens of the invention can be used in immunoassays to detect antibody levels (or, conversely, anti-Neisserial antibodies can be used to detect antigen levels). Immunoassays based on well defined, recombinant antigens can be developed to replace invasive diagnostics methods. Antibodies to Neisserial proteins within biological samples, including for example, blood or serum samples, can be detected. Design of the immunoassays is subject to a great deal of variation, and a variety of these are known in the art. Protocols for the immunoassay may be based, for example, upon competition, or direct reaction, or sandwich type assays. Protocols may also, for example, use solid supports, or may be by immunoprecipitation. Most assays involve the use of labeled antibody or polypeptide; the labels may be, for example, fluorescent, chemiluminescent, radioactive, or dye molecules. Assays which amplify the signals from the probe are also known; examples of which are assays which utilize biotin and avidin, and enzyme-labeled and mediated immunoassays, such as ELISA assays.

Kits suitable for immunodiagnosis and containing the appropriate labeled reagents are constructed by packaging the appropriate materials, including the compositions of the invention, in suitable

containers, along with the remaining reagents and materials (for example, suitable buffers, salt solutions, *etc.*) required for the conduct of the assay, as well as suitable set of assay instructions.

### Nucleic Acid Hybridisation

“Hybridization” refers to the association of two nucleic acid sequences to one another by hydrogen bonding. Typically, one sequence will be fixed to a solid support and the other will be free in solution. Then, the two sequences will be placed in contact with one another under conditions that favor hydrogen bonding. Factors that affect this bonding include: the type and volume of solvent; reaction temperature; time of hybridization; agitation; agents to block the non-specific attachment of the liquid phase sequence to the solid support (Denhardt's reagent or BLOTTO); concentration of the sequences; use of compounds to increase the rate of association of sequences (dextran sulfate or polyethylene glycol); and the stringency of the washing conditions following hybridization. See Sambrook *et al.* [*supra*] Volume 2, chapter 9, pages 9.47 to 9.57.

“Stringency” refers to conditions in a hybridization reaction that favor association of very similar sequences over sequences that differ. For example, the combination of temperature and salt concentration should be chosen that is approximately 120 to 200°C below the calculated  $T_m$  of the hybrid under study. The temperature and salt conditions can often be determined empirically in preliminary experiments in which samples of genomic DNA immobilized on filters are hybridized to the sequence of interest and then washed under conditions of different stringencies. See Sambrook *et al.* at page 9.50.

Variables to consider when performing, for example, a Southern blot are (1) the complexity of the DNA being blotted and (2) the homology between the probe and the sequences being detected. The total amount of the fragment(s) to be studied can vary a magnitude of 10, from 0.1 to 1 µg for a plasmid or phage digest to  $10^{-9}$  to  $10^{-8}$  g for a single copy gene in a highly complex eukaryotic genome. For lower complexity polynucleotides, substantially shorter blotting, hybridization, and exposure times, a smaller amount of starting polynucleotides, and lower specific activity of probes can be used. For example, a single-copy yeast gene can be detected with an exposure time of only 1 hour starting with 1 µg of yeast DNA, blotting for two hours, and hybridizing for 4-8 hours with a probe of  $10^8$  cpm/µg. For a single-copy mammalian gene a conservative approach would start with 10 µg of DNA, blot overnight, and hybridize overnight in the presence of 10% dextran sulfate using a probe of greater than  $10^8$  cpm/µg, resulting in an exposure time of ~24 hours.

Several factors can affect the melting temperature ( $T_m$ ) of a DNA-DNA hybrid between the probe and the fragment of interest, and consequently, the appropriate conditions for hybridization and washing. In many cases the probe is not 100% homologous to the fragment. Other commonly encountered variables include the length and total G+C content of the hybridizing sequences and the ionic strength and formamide content of the hybridization buffer. The effects of all of these factors can be approximated by a single equation:

$$T_m = 81 + 16.6(\log_{10} C_i) + 0.4[\%(G + C)] - 0.6(\%\text{formamide}) - 600/n - 1.5(\%\text{mismatch}).$$

where  $C_i$  is the salt concentration (monovalent ions) and  $n$  is the length of the hybrid in base pairs (slightly modified from Meinkoth & Wahl (1984) *Anal. Biochem.* 138: 267-284).

- 10 In designing a hybridization experiment, some factors affecting nucleic acid hybridization can be conveniently altered. The temperature of the hybridization and washes and the salt concentration during the washes are the simplest to adjust. As the temperature of the hybridization increases (*ie.* stringency), it becomes less likely for hybridization to occur between strands that are nonhomologous, and as a result, background decreases. If the radiolabeled probe is not completely
- 15 homologous with the immobilized fragment (as is frequently the case in gene family and interspecies hybridization experiments), the hybridization temperature must be reduced, and background will increase. The temperature of the washes affects the intensity of the hybridizing band and the degree of background in a similar manner. The stringency of the washes is also increased with decreasing salt concentrations.
- 20 In general, convenient hybridization temperatures in the presence of 50% formamide are 42°C for a probe with is 95% to 100% homologous to the target fragment, 37°C for 90% to 95% homology, and 32°C for 85% to 90% homology. For lower homologies, formamide content should be lowered and temperature adjusted accordingly, using the equation above. If the homology between the probe and the target fragment are not known, the simplest approach is to start with both hybridization and
- 25 wash conditions which are nonstringent. If non-specific bands or high background are observed after autoradiography, the filter can be washed at high stringency and reexposed. If the time required for exposure makes this approach impractical, several hybridization and/or washing stringencies should be tested in parallel.

Nucleic Acid Probe Assays

Methods such as PCR, branched DNA probe assays, or blotting techniques utilizing nucleic acid probes according to the invention can determine the presence of cDNA or mRNA. A probe is said to "hybridize" with a sequence of the invention if it can form a duplex or double stranded complex, which is stable enough to be detected.

The nucleic acid probes will hybridize to the Neisserial nucleotide sequences of the invention (including both sense and antisense strands). Though many different nucleotide sequences will encode the amino acid sequence, the native Neisserial sequence is preferred because it is the actual sequence present in cells. mRNA represents a coding sequence and so a probe should be complementary to the coding sequence; single-stranded cDNA is complementary to mRNA, and so a cDNA probe should be complementary to the non-coding sequence.

The probe sequence need not be identical to the Neisserial sequence (or its complement) — some variation in the sequence and length can lead to increased assay sensitivity if the nucleic acid probe can form a duplex with target nucleotides, which can be detected. Also, the nucleic acid probe can include additional nucleotides to stabilize the formed duplex. Additional Neisserial sequence may also be helpful as a label to detect the formed duplex. For example, a non-complementary nucleotide sequence may be attached to the 5' end of the probe, with the remainder of the probe sequence being complementary to a Neisserial sequence. Alternatively, non-complementary bases or longer sequences can be interspersed into the probe, provided that the probe sequence has sufficient complementarity with the a Neisserial sequence in order to hybridize therewith and thereby form a duplex which can be detected.

The exact length and sequence of the probe will depend on the hybridization conditions, such as temperature, salt condition and the like. For example, for diagnostic applications, depending on the complexity of the analyte sequence, the nucleic acid probe typically contains at least 10-20 nucleotides, preferably 15-25, and more preferably at least 30 nucleotides, although it may be shorter than this. Short primers generally require cooler temperatures to form sufficiently stable hybrid complexes with the template.

Probes may be produced by synthetic procedures, such as the triester method of Matteucci *et al.* [*J. Am. Chem. Soc.* (1981) 103:3185], or according to Urdea *et al.* [*Proc. Natl. Acad. Sci. USA* (1983) 80: 7461], or using commercially available automated oligonucleotide synthesizers.

The chemical nature of the probe can be selected according to preference. For certain applications, DNA or RNA are appropriate. For other applications, modifications may be incorporated *eg.* backbone modifications, such as phosphorothioates or methylphosphonates, can be used to increase *in vivo* half-life, alter RNA affinity, increase nuclease resistance *etc.* [*eg.* see Agrawal & Iyer (1995) *Curr Opin Biotechnol* 6:12-19; Agrawal (1996) *TIBTECH* 14:376-387]; analogues such as peptide nucleic acids may also be used [*eg.* see Corey (1997) *TIBTECH* 15:224-229; Buchardt *et al.* (1993) *TIBTECH* 11:384-386].

Alternatively, the polymerase chain reaction (PCR) is another well-known means for detecting small amounts of target nucleic acids. The assay is described in: Mullis *et al.* [*Meth. Enzymol.* (1987) 155: 335-350]; US patents 4,683,195 and 4,683,202. Two "primer" nucleotides hybridize with the target nucleic acids and are used to prime the reaction. The primers can comprise sequence that does not hybridize to the sequence of the amplification target (or its complement) to aid with duplex stability or, for example, to incorporate a convenient restriction site. Typically, such sequence will flank the desired Neisserial sequence.

A thermostable polymerase creates copies of target nucleic acids from the primers using the original target nucleic acids as a template. After a threshold amount of target nucleic acids are generated by the polymerase, they can be detected by more traditional methods, such as Southern blots. When using the Southern blot method, the labelled probe will hybridize to the Neisserial sequence (or its complement).

Also, mRNA or cDNA can be detected by traditional blotting techniques described in Sambrook *et al* [*supra*]. mRNA, or cDNA generated from mRNA using a polymerase enzyme, can be purified and separated using gel electrophoresis. The nucleic acids on the gel are then blotted onto a solid support, such as nitrocellulose. The solid support is exposed to a labelled probe and then washed to remove any unhybridized probe. Next, the duplexes containing the labeled probe are detected. Typically, the probe is labelled with a radioactive moiety.

## BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1-20 show biochemical data obtained in the Examples, and also sequence analysis, for ORFs 37, 5, 2, 15, 22, 28, 32, 4, 61, 76, 89, 97, 106, 138, 23, 25, 27, 79, 85 and 132. M1 and M2 are molecular weight markers. Arrows indicate the position of the main recombinant product or, in Western blots, the position of the main *N.meningitidis* immunoreactive band. TP indicates *N.meningitidis* total protein extract; OMV indicates *N.meningitidis* outer membrane vesicle preparation. In bactericidal assay results: a diamond (♦) shows preimmune data; a triangle (▲) shows GST control data; a circle (●) shows data with recombinant *N.meningitidis* protein. Computer analyses show a hydrophilicity plot (upper), an antigenic index plot (middle), and an AMPHI analysis (lower). The AMPHI program has been used to predict T-cell epitopes [Gao *et al.* (1989) *J. Immunol.* 143:3007; Roberts *et al.* (1996) *AIDS Res Hum Retrovir* 12:593; Quakyi *et al.* (1992) *Scand J Immunol* suppl.11:9) and is available in the Protean package of DNASTAR, Inc. (1228 South Park Street, Madison, Wisconsin 53715 USA).

Figure 21 shows an alignment comparison of amino acid sequences for ORF 4 for several strains of *Neisseria*. Dark shading indicates regions of homology, and gray shading indicates the conservation of amino acids with similar characteristics. The Figure demonstrates a high degree of conservation among the various strains, further confirming its utility as an antigen for both vaccines and diagnostics.

## EXAMPLES

The examples describe nucleic acid sequences which have been identified in *N.meningitidis*, along with their putative translation products, and also those of *N.gonorrhoeae*. Not all of the nucleic acid sequences are complete *ie.* they encode less than the full-length wild-type protein.

The examples are generally in the following format:

- a nucleotide sequence which has been identified in *N.meningitidis* (strain B)
- the putative translation product of this sequence
- a computer analysis of the translation product based on database comparisons
- corresponding gene and protein sequences identified in *N.meningitidis* (strain A) and in *N.gonorrhoeae*

- a description of the characteristics of the proteins which indicates that they might be suitably antigenic
- results of biochemical analysis (expression, purification, ELISA, FACS *etc.*)

5 The examples typically include details of sequence identity between species and strains. Proteins that are similar in sequence are generally similar in both structure and function, and the sequence identity often indicates a common evolutionary origin. Comparison with sequences of proteins of known function is widely used as a guide for the assignment of putative protein function to a new sequence and has proved particularly useful in whole-genome analyses.

10 Sequence comparisons were performed at NCBI (<http://www.ncbi.nlm.nih.gov>) using the algorithms BLAST, BLAST2, BLASTn, BLASTp, tBLASTn, BLASTx, & tBLASTx [*eg.* see also Altschul *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25:2289-3402]. Searches were performed against the following databases: non-redundant GenBank+EMBL+DDBJ+PDB sequences and non-redundant GenBank CDS translations+PDB+SwissProt+SPupdate+PIR sequences.

15 To compare Meningococcal and Gonococcal sequences, the tBLASTx algorithm was used, as implemented at [http://www.genome.ou.edu/gono\\_blast.html](http://www.genome.ou.edu/gono_blast.html). The FASTA algorithm was also used to compare the ORFs (from GCG Wisconsin Package, version 9.0).

20 Dots within nucleotide sequences (*eg.* position 495 in SEQ ID 11) represent nucleotides which have been arbitrarily introduced in order to maintain a reading frame. In the same way, double-underlined nucleotides were removed. Lower case letters (*eg.* position 496 in SEQ ID 11) represent ambiguities which arose during alignment of independent sequencing reactions (some of the nucleotide sequences in the examples are derived from combining the results of two or more experiments).

25 Nucleotide sequences were scanned in all six reading frames to predict the presence of hydrophobic domains using an algorithm based on the statistical studies of Esposti *et al.* [Critical evaluation of the hydropathy of membrane proteins (1990) *Eur J Biochem* 190:207-219]. These domains represent potential transmembrane regions or hydrophobic leader sequences.



Open reading frames were predicted from fragmented nucleotide sequences using the program ORFFINDER (NCBI).

Underlined amino acid sequences indicate possible transmembrane domains or leader sequences in the ORFs, as predicted by the PSORT algorithm (<http://www.psорт.nibb.ac.jp>). Functional domains were also predicted using the MOTIFS program (GCG Wisconsin & PROSITE).

Various tests can be used to assess the *in vivo* immunogenicity of the proteins identified in the examples. For example, the proteins can be expressed recombinantly and used to screen patient sera by immunoblot. A positive reaction between the protein and patient serum indicates that the patient has previously mounted an immune response to the protein in question *ie.* the protein is an immunogen. This method can also be used to identify immunodominant proteins.

The recombinant protein can also be conveniently used to prepare antibodies *eg.* in a mouse. These can be used for direct confirmation that a protein is located on the cell-surface. Labelled antibody (*eg.* fluorescent labelling for FACS) can be incubated with intact bacteria and the presence of label on the bacterial surface confirms the location of the protein.

In particular, the following methods (A) to (S) were used to express, purify and biochemically characterise the proteins of the invention:

#### A) Chromosomal DNA preparation

*N.meningitidis* strain 2996 was grown to exponential phase in 100ml of GC medium, harvested by centrifugation, and resuspended in 5ml buffer (20% Sucrose, 50mM Tris-HCl, 50mM EDTA, pH8). After 10 minutes incubation on ice, the bacteria were lysed by adding 10ml lysis solution (50mM NaCl, 1% Na-Sarkosyl, 50µg/ml Proteinase K), and the suspension was incubated at 37°C for 2 hours. Two phenol extractions (equilibrated to pH 8) and one  $\text{ChCl}_3$ /isoamylalcohol (24:1) extraction were performed. DNA was precipitated by addition of 0.3M sodium acetate and 2 volumes ethanol, and was collected by centrifugation. The pellet was washed once with 70% ethanol and redissolved in 4ml buffer (10mM Tris-HCl, 1mM EDTA, pH 8). The DNA concentration was measured by reading the OD at 260 nm.

**B) Oligonucleotide design**

Synthetic oligonucleotide primers were designed on the basis of the coding sequence of each ORF, using (a) the meningococcus B sequence when available, or (b) the gonococcus/meningococcus A sequence, adapted to the codon preference usage of meningococcus as necessary. Any predicted  
 5 signal peptides were omitted, by deducing the 5'-end amplification primer sequence immediately downstream from the predicted leader sequence.

For most ORFs, the 5' primers included two restriction enzyme recognition sites (*Bam*HI-*Nde*I, *Bam*HI-*Nhe*I, or *Eco*RI-*Nhe*I, depending on the gene's own restriction pattern); the 3' primers included a *Xho*I restriction site. This procedure was established in order to direct the cloning of  
 10 each amplification product (corresponding to each ORF) into two different expression systems: pGEX-KG (using either *Bam*HI-*Xho*I or *Eco*RI-*Xho*I), and pET21b+ (using either *Nde*I-*Xho*I or *Nhe*I-*Xho*I).

	5'-end primer tail:	<u>CGCGGATCCCATATG</u>	( <i>Bam</i> HI- <i>Nde</i> I)
		<u>CGCGGATCCGCTAGC</u>	( <i>Bam</i> HI- <i>Nhe</i> I)
15		<u>CCGGAATTCTAGCTAGC</u>	( <i>Eco</i> RI- <i>Nhe</i> I)
	3'-end primer tail:	<u>CCCGCTCGAG</u>	( <i>Xho</i> I)

For ORFs 5, 15, 17, 19, 20, 22, 27, 28, 65 & 89, two different amplifications were performed to clone each ORF in the two expression systems. Two different 5' primers were used for each ORF; the same 3' *Xho*I primer was used as before:

20	5'-end primer tail:	<u>GGAATTCCATATGGCCATGG</u>	( <i>Nde</i> I)
	5'-end primer tail:	<u>CGGGATCC</u>	( <i>Bam</i> HI)

ORF 76 was cloned in the pTRC expression vector and expressed as an amino-terminus His-tag fusion. In this particular case, the predicted signal peptide was included in the final product. *Nhe*I-*Bam*HI restriction sites were incorporated using primers:

25	5'-end primer tail:	<u>GATCAGCTAGCCATATG</u>	( <i>Nhe</i> I)
	3'-end primer tail:	<u>CGGGATCC</u>	( <i>Bam</i> HI)

As well as containing the restriction enzyme recognition sequences, the primers included nucleotides which hybridized to the sequence to be amplified. The number of hybridizing nucleotides depended on the melting temperature of the whole primer, and was determined for each primer using the formulae:

$$T_m = 4 (G+C) + 2 (A+T) \quad (\text{tail excluded})$$

$$T_m = 64.9 + 0.41 (\% \text{ GC}) - 600/N \quad (\text{whole primer})$$

The average melting temperature of the selected oligos were 65-70°C for the whole oligo and 50-55°C for the hybridising region alone.

**Table I** shows the forward and reverse primers used for each amplification. In certain cases, it will be noted that the sequence of the primer does not exactly match the sequence in the ORF. When initial amplifications were performed, the complete 5' and/or 3' sequence was not known for some meningococcal ORFs, although the corresponding sequences had been identified in gonococcus. For amplification, the gonococcal sequences could thus be used as the basis for primer design, altered to take account of codon preference. In particular, the following codons were changed: ATA→ATT; TCG→TCT; CAG→CAA; AAG→AAA; GAG→GAA; CGA→CGC; CGG→CGC; GGG→GGC. Italicised nucleotides in Table I indicate such a change. It will be appreciated that, once the complete sequence has been identified, this approach is generally no longer necessary.

**TABLE I – PCR primers**

ORF	Primer	Sequence	Restriction sites
<b>ORF 1</b>	Forward	CGCGGATCCGCTAGC-GGACACACTTATTTTCGG <SEQ ID 924>	BamHI-NheI
	Reverse	CCC <u>GCTCGAG</u> -CCAGCGGTAGCCTAATT <SEQ ID 925>	XhoI
<b>ORF 2</b>	Forward	GCGGATCCCATATG-TTTGATTTCGGTTTGGG <SEQ ID 926>	BamHI-NdeI
	Reverse	CCC <u>GCTCGAG</u> -GACGGCATAACGGCG <SEQ ID 927>	XhoI
<b>ORF 2-1</b>	Forward	GCGGATCCCATATG-TTTGATTTCGGTTTGGG <SEQ ID 928>	BamHI-NdeI
	Reverse	CCC <u>GCTCGAG</u> -TGATTTACGGACGCGCA <SEQ ID 929>	XhoI
<b>ORF 4</b>	Forward	GCGGATCCCATATG-TGCGGAGGTCAAAAAGAC <SEQ ID 930>	BamHI-NdeI
	Reverse	CCC <u>GCTCGAG</u> -TTTGGCTGCGCCTTC <SEQ ID 931>	XhoI

ORF 5	Forward	GGAATTCATATGGCCATGG-TGGAAGGCGCACAACC <SEQ ID 932>	NdeI-NcoI
	Forward	CGGGATCC-ATGGAAGGCGCACAAC <SEQ ID 933>	BamHI
	Reverse	CCCGCTCGAG-GACTGTGCAAAAACGG <SEQ ID 934>	XhoI
ORF 6	Forward	CGCGGATCCCATATG-ACCCGTCAATCTCTGCA <SEQ ID 935>	BamHI-NdeI
	Reverse	CCCGCTCGAG-TGCGCCGAACACTTTC <SEQ ID 936>	XhoI
ORF 7	Forward	CGCGGATCCGCTAGC-GCGCTGCTTTTGTTC <SEQ ID 937>	BamHI-NheI
	Reverse	CCCGCTCGAG-TTTCAAAATATATTGCGGA <SEQ ID 938>	XhoI
ORF 8	Forward	GCGGATCCCATATG-GCTCAACTGCTTCGTAC <SEQ ID 939>	BamHI-NdeI
	Reverse	CCCGCTCGAG-AGCAGGCTTGGCGC <SEQ ID 940>	XhoI
ORF 9	Forward	CGCGGATCCCATATG-CCGAAGGAAGTCGAAA <SEQ ID 941>	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTTCGAGGTTTTCGGG <SEQ ID 942>	XhoI
ORF 10	Forward	GCGGATCCCATATG-GACACAAAAGAAATCCTC <SEQ ID 943>	BamHI-NdeI
	Reverse	CCCGCTCGAG-TAATGGGAAACCTTGTTTT <SEQ ID 944>	XhoI
ORF 11	Forward	GCGGATCCCATATG-GCGGTCAACCTCTACG <SEQ ID 945>	BamHI-NdeI
	Reverse	CCCGCTCGAG-GGAAACGACTTCGCC <SEQ ID 946>	XhoI
ORF 13	Forward	CGCGGATCCCATATG-GCTCTGCTTCCGCGC <SEQ ID 947>	BamHI-NdeI
	Reverse	CCCGCTCGAG-AGGGTGTGTGATAATAAG <SEQ ID 948>	XhoI
ORF 15	Forward	GGAATTCATATGGCCATGG-GCGGGACACTGACAG <SEQ ID 949>	NdeI-NcoI
	Forward	CGGGATCC-TGCGGGACACTGACAGG <SEQ ID 950>	BamHI
	Reverse	CCCGCTCGAG-AGGTTGCCTTGTCTATG <SEQ ID 951>	XhoI
ORF 17	Forward	GGAATTCATATGGCCATGG-TTGCCGGCCTGTTTCG <SEQ ID 952>	NdeI-NcoI
	Forward	CGGGATCC-ATTGCCGGCCTGTTTCG <SEQ ID 953>	BamHI
	Reverse	CCCGCTCGAG-AAGCAGGTTGTACAGC <SEQ ID 954>	XhoI
ORF 18	Forward	GCGGATCCCATATG-ATTTTGCTGCATTGGAT <SEQ ID 955>	BamHI-NdeI
	Reverse	CCCGCTCGAG-TCTTCCAATTTCTGAAAGC <SEQ ID 956>	XhoI
ORF 19	Forward	GGAATTCATATGGCCATGG-TCGCCAGTGTTTTTACC <SEQ ID 957>	NdeI-NcoI
	Forward	CGGGATCC-TTCGCCAGTGTTTTTACCG <SEQ ID 958>	BamHI
	Reverse	CCCGCTCGAG-GGTGTTTTTGAAGCTGCC <SEQ ID 959>	XhoI
ORF 20	Forward	GGAATTCATATGGCCATGG-TCGGCGCGGGTATG <SEQ ID 960>	NdeI-NcoI

600410 54500650

	Forward	CGGGATCC-TTCGGCGCGGGTATG <SEQ ID 961>	BamHI
	Reverse	CCCGCTCGAG-CGGCGAGCGAGAGCA <SEQ ID 962>	XhoI
ORF 22	Forward	GGAATTCCATATGGCCATGG-TGATTAAATCAAAAAAGGTCT <SEQ ID 963>	NdeI-NcoI
	Forward	CGGGATCC-ATGATTAAATCAAAAAAGGTCTAAACC <SEQ ID 964>	BamHI
	Reverse	CCCGCTCGAG-ATTATGATAGCGGCC <SEQ ID 965>	XhoI
ORF 23	Forward	CGCGGATCCCATATG-GATGTTTCTGTTTCAGAC <SEQ ID 966>	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTTAAACCGATAGGTAAACG <SEQ ID 967>	XhoI
ORF 24	Forward	GGAATTCCATATGGCCATGG -TGATGCCGGAATGGTG <SEQ ID 968>	NdeI-NcoI
	Forward	CGGGATCC-ATGATGCCGGAATGGTG <SEQ ID 969>	BamHI
	Reverse	CCCGCTCGAG-TGTCAGCGTGCGCA <SEQ ID 970>	XhoI
ORF 25	Forward	GCGGATCCCATATG-TATCGCAAATGATTGC <SEQ ID 971>	BamHI-NdeI
	Reverse	CCCGCTCGAG-ATCGATGGAATAGCCG <SEQ ID 972>	XhoI
ORF 26	Forward	GCGGATCCCATATG -CAGCTGATCGACTATTC <SEQ ID 973>	BamHI-NdeI
	Reverse	CCCGCTCGAG-GACATCGGCGGTTTT <SEQ ID 974>	XhoI
ORF 27	Forward	GGAATTCCATATGGCCATGG-AGACCTATTCTGTTTA <SEQ ID 974>	NdeI-NcoI
	Forward	CGGGATCC- CAGACCTATTCTGTTATTTTAATC <SEQ ID 975>	BamHI
	Reverse	CCCGCTCGAG-GGGTTCGATTAAATAACCAT <SEQ ID 976>	XhoI
ORF 28	Forward	GGAATTCCATATGGCCATGG-ACGGCTGTACGTTGATGT <SEQ ID 977>	NdeI-NcoI
	Forward	CGGGATCC-AACGGCTGTACGTTGATG <SEQ ID 978>	BamHI
	Reverse	CCCGCTCGAG-TTTGTGAGGAATTCGCG <SEQ ID 979>	XhoI
ORF 29	Forward	GCGGATCCCATATG -AACGGTTTGATGCCCCG <SEQ ID 980>	BamHI-NdeI
	Forward	CGCGGATCCGCTAGC-AACGGTTTGATGCCCCG <SEQ ID 981>	BamHI-NheI
	Reverse	CCCGCTCGAG-TTTGTCTAAGTTCCTGATATG <SEQ ID 982>	XhoI
ORF 32	Forward	CGCGGATCCCATATG-AATACTCCTCCTTTG <SEQ ID 983>	BamHI-NdeI
	Reverse	CCCGCTCGAG-GCGTATTTTTGATGCTTTG <SEQ ID 984>	XhoI
ORF 33	Forward	GCGGATCCCATATG -ATTGATAGGGATCGTATG <SEQ ID 985>	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTGATCTTCAAACGGCC <SEQ ID 986>	XhoI
ORF 35	Forward	GCGGATCCCATATG-TTCAGAGCTCAGCTT <SEQ ID 987>	BamHI-NdeI
	Forward	CGCGGATCCGCTAGC-TTCAGAGCTCAGCTT <SEQ ID 988>	BamHI-NheI
	Reverse	CCCGCTCGAG-AAACAGCCATTTGAGCGA <SEQ ID 989>	XhoI

<b>ORF 37</b>	Forward	GCGGATCCCATATG-GATGACGTATCGGATTTT <SEQ ID 990>	BamHI-NdeI
	Reverse	CCCGCTCGAG-ATAGCCCGCTTTTCAGG <SEQ ID 991>	XhoI
<b>ORF 58</b>	Forward	CGCGGATCCGCTAGC-TCCGAACGCGAGTGGAT <SEQ ID 992>	BamHI-NheI
	Reverse	CCCGCTCGAG-AGCATTGTCCAAGGGGAC <SEQ ID 993>	XhoI
<b>ORF 65</b>	Forward	GGAATTCCATATGGCCATGG -TGCTGTATCTGAATCAAG <SEQ ID 994>	NdeI-NcoI
	Forward	CGGGATCC-TTGCTGTATCTGAATCAAGG <SEQ ID 995>	BamHI
	Reverse	CCCGCTCGAG-CCGCATCGGCAGACA <SEQ ID 996>	XhoI
<b>ORF 66</b>	Forward	GCGGATCCCATATG-TACGCATTTACCGCCG <SEQ ID 997>	BamHI-NdeI
	Reverse	CCCGCTCGAG-TGGATTTGCAGAGATGG <SEQ ID 998>	XhoI
<b>ORF 72</b>	Forward	CGCGGATCCCATATG- AATGCAGTAAAAATATCTGA <SEQ ID 999>	BamHI-NdeI
	Reverse	CCCGCTCGAG-GCCTGAGACCTTTGCAA <SEQ ID 1000>	XhoI
<b>ORF 73</b>	Forward	GCGGATCCCATATG-AGATTTTTCGGTATCGG <SEQ ID 1001>	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTCATCTTTTTCATGTTTCG <SEQ ID 1002>	XhoI
<b>ORF 75</b>	Forward	GCGGATCCCATATG- TCTGTCTTTCAAACGGC <SEQ ID 1003>	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTGTTTTTGCAAGACAG <SEQ ID 1004>	XhoI
<b>ORF 76</b>	Forward	GATCAGCTAGCCATATG-AAACAGAAAAAACCGC <SEQ ID 1005>	NheI-NdeI
	Reverse	CGGGATCC-TTACGGTTTGACACCGTT <SEQ ID 1006>	BamHI
<b>ORF 79</b>	Forward	CGCGGATCCCATATG-GTTTCGCGCCCG <SEQ ID 1007>	BamHI-NdeI
	Reverse	CCCGCTCGAG-GTGCTGATGCGCTTCG <SEQ ID 1008>	XhoI
<b>ORF 83</b>	Forward	GCGGATCCCATATG-AAAACCCTGCTGCTGC <SEQ ID 1009>	BamHI-NdeI
	Reverse	CCCGCTCGAG-GCCGCCTTTGCGGC <SEQ ID 1010>	XhoI
<b>ORF 84</b>	Forward	GCGGATCCCATATG-GCAGAGATCTGTTTG <SEQ ID 1011>	BamHI-NdeI
	Reverse	CCCGCTCGAG-GTTTGCCGATCCGACCA <SEQ ID 1012>	XhoI
<b>ORF 85</b>	Forward	CGCGGATCCCATATG- GCGGTTTGGGGCGGA <SEQ ID 1013>	BamHI-NdeI
	Reverse	CCCGCTCGAG-TCGGCGCGGCGGGC <SEQ ID 1014>	XhoI
<b>ORF 89</b>	Forward	GGAATTCCATATGGCCATGG-CCATACCTTCTTATCA <SEQ ID 1015>	NdeI-NcoI
	Forward	CGGGATCC-GCCATACCTTCTTATCAGAG <SEQ ID 1016>	BamHI
	Reverse	CCCGCTCGAG-TTTTTTGCGATTAGAAAAAGC <SEQ ID 1017>	XhoI

<b>ORF 97</b>	Forward	GCGGATCCCATATG-CATCCTGCCAGCGAAC <SEQ ID 1018>	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-TTCGCCTACGGTTTTTTTG <SEQ ID 1019>	
<b>ORF 98</b>	Forward	GCGGATCCCATATG-ACGGTAACTGCGG <SEQ ID 1020>	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-TTGTGTTCGGGCAAATC <SEQ ID 1021>	
<b>ORF 100</b>	Forward	GCGGATCCCATATG-TCGGGCATTACACCG <SEQ ID 1022>	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-ACGGGTTCGGCGGAA <SEQ ID 1023>	
<b>ORF 101</b>	Forward	GCGGATCCCATATG-ATTATCAAAGAAACCTC <SEQ ID 1024>	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-TTTTCCGCCTTTCAATGT <SEQ ID 1025>	
<b>ORF 102</b>	Forward	GCGGATCCCATATG-GCAGGGCTGTTTTACC <SEQ ID 1026>	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-AAACGGTTGAACACGAC <SEQ ID 1027>	
<b>ORF 103</b>	Forward	GCGGATCCCATATG-AACCACGACATCAC <SEQ ID 1028>	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-CAGCCACAGGACGGC <SEQ ID 1029>	
<b>ORF 104</b>	Forward	GCGGATCCCATATG-ACGTGGGGAACGC <SEQ ID 1030>	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-GCGGCGTTTGAACGGC <SEQ ID 1031>	
<b>ORF 105</b>	Forward	GCGGATCCCATATG-ACCAAATTTCAAACCCCTC <SEQ ID 1032>	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-TAAACGAATGCCGTCCAG <SEQ ID 1033>	
<b>ORF 106</b>	Forward	GCGGATCCCATATG-AGGATAACCGACGGCG <SEQ ID 1034>	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-TTTGTTCCCGATGATGTT <SEQ ID 1035>	
<b>ORF 109</b>	Forward	GCGGATCCCATATG-GAAGATTTATATATAATACTCG <SEQ ID 1036>	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-ATCAGCTTCGAACCGAAG <SEQ ID 1037>	
<b>ORF110</b>	Forward	AAAGAATTC-ATGAGTAAATCCCGTAGATCTCCC <SEQ ID 1038>	EcoRI
	Reverse	AAACTGCAG-GGAAACCACATCCGCACTCTGCC <SEQ ID 1039>	PstI
<b>ORF111</b>	Forward	AAAGAATTC-GCACC GCAAAAGGCAAAAACCGCA <SEQ ID 1040>	EcoRI
	Reverse	AAACTGCAG-TCTGCGCGTTTCGGGCAGGGTGG <SEQ ID 1041>	PstI
<b>ORF113</b>	Forward	AAAGAATTC-ATGAACAAAACCTCTATCGTGTGATTTTCAACCG <SEQ ID 1042>	EcoRI
	Reverse	AAACTGCAG-TTACGAATGCCTGCTTGCTCGACCGTACTG <SEQ ID 1043>	PstI
<b>ORF115</b>	Forward	AAAGAATTC-TTGCTTGTGCAACAGAAAAGACGG <SEQ ID 1044>	EcoRI
	Reverse	AAAAAAGTCGAC-CTATTTTTAGGGGCTTTGCTTGTGTTGAAAAGCCTGCC <SEQ ID 1045>	SalI

60640-8730000

<b>ORF119</b>	Forward	AAAGAATTC-TACAACATGTATCAGGAAAACCAATACCG <SEQ ID 1046>	EcoRI
	Reverse	AAACTGCAG-TTATGAAAACAGGCGCAGGGCGGTTTTGCC <SEQ ID 1047>	PstI
<b>ORF120</b>	Forward	AAAGAATTC-GCAAGGCTACCCCAATCCGCCGTG <SEQ ID 1048>	EcoRI
	Reverse	AAACTGCAG-CGGTTTGCTGCCTGGCCGTGAT <SEQ ID 1049>	PstI
<b>ORF121</b>	Forward	AAAGAATTC-GCCTTGGTCTGGCTGGTTTTTCGC <SEQ ID 1050>	EcoRI
	Reverse	AAACTGCAG-TCATCCGCCACCCACCTCGCCATCCATC <SEQ ID 1051>	PstI
<b>ORF122</b>	Forward	AAAAAAGTCGAC-ATGTCTTACCGCGCAAGCAGTTCTCC <SEQ ID 1052>	Sall
	Reverse	AAACTGCAG-TCAGGAACACAAACGATGACGAATATCCGTATC <SEQ ID 1053>	PstI
<b>ORF125</b>	Forward	AAAGAATTC-GCGCTGTTTTTTGCGGCGGCGTAT <SEQ ID 1054>	EcoRI
	Reverse	AAACTGCAG-CGCCGTTTCAAGACGAAAAAGTCG <SEQ ID 1055>	PstI
<b>ORF126</b>	Forward	AAAGAATTC-GCGGAAACGGTCTGAAG <SEQ ID 1056>	EcoRI
	Reverse	AAACTGCAG-TTAATCTGTCTTCCGATATAC <SEQ ID 1057>	PstI
<b>ORF127</b>	Forward	AAAGAATTC-ATGACTGATAATCGGGGGTTTACG <SEQ ID 1058>	EcoRI
	Reverse	AAAAAAGTCGAC-CTTAAGTAACTGCAGTCCTTATC <SEQ ID 1059>	Sall
<b>ORF128</b>	Forward	AAAGAATTC-ATGCAAGCTGTCCGCTACAGGCC <SEQ ID 1060>	EcoRI
	Reverse	AAACTGCAG-CTATTGCAATGCGCCGCCCGGGAATGTTGAGCAGGCG <SEQ ID 1061>	PstI
<b>ORF129</b>	Forward	AAAGAATTC-ATGGATTTTCGTTTGTGACATTATTACGAATACCG <SEQ ID 1062>	EcoRI
	Reverse	AAACTGCAG-TTATTTTTTGATGAAATTTTGGGGCGG <SEQ ID 1063>	PstI
<b>ORF130</b>	Forward	AAAGAATTC-GCAGTACTTGCCATTCTCGGTGCG <SEQ ID 1064>	EcoRI
	Reverse	AAACTGCAG-CTCCGGATCGTCTGTAAACGCATT <SEQ ID 1065>	PstI
<b>ORF 131</b>	Forward	GCGGATCCCATATG-GAAATTCGGGCAATAAAAT <SEQ ID 1066>	BamHI-NdeI
	Reverse	CCCGCTCGAG-CCAGCGGACGCGTTC <SEQ ID 1067>	XhoI
<b>ORF 132</b>	Forward	GCGGATCCCATATG-AAAGAAGCGGGGTTTG <SEQ ID 1068>	BamHI-NdeI
	Reverse	CCCGCTCGAG-CCAATCTGCCAGCCGT <SEQ ID 1069>	XhoI
<b>ORF 133</b>	Forward	CGCGGATCCCATATG-GAAGATGCAGGGCGCG <SEQ ID 1070>	BamHI-NdeI

60640-673000



<b>ORF 134</b>	Reverse	<u>CCCGCTCGAG</u> -AACTTGTAGCTCATCGT <SEQ ID 1071>	XhoI
	Forward	GCGGATCCCATATG-TCTGTGCAAGCAGTATTG <SEQ ID 1072>	BamHI-NdeI
	Reverse	<u>CCCGCTCGAG</u> -ATCCTGTGCCAATGCG <SEQ ID 1073>	XhoI
<b>ORF 135</b>	Forward	GCGGATCCCATATG-CCGTCTGAAAAAGCTTT <SEQ ID 1074>	BamHI-NdeI
	Reverse	<u>CCCGCTCGAG</u> -AAATACCGCTGAGGATG <SEQ ID 1075>	XhoI
<b>ORF 136</b>	Forward	GCGGATCCGCTAGC-ATGAAGCGGCGTATAGCC <SEQ ID 1076>	BamHI-NheI
	Reverse	<u>CCCGCTCGAG</u> -TTCCGAATATTTGGAACCTTT <SEQ ID 1077>	XhoI
<b>ORF 137</b>	Forward	GCGGATCCCATATG-GGCACGGCGGGAAATA <SEQ ID 1078>	BamHI-NdeI
	Reverse	<u>CCCGCTCGAG</u> -ATAACGGTATGCCGCC <SEQ ID 1079>	XhoI
<b>ORF 138</b>	Forward	GCGGATCCCATATG-TTTCGTTTACAATTCAGGC <SEQ ID 1080>	BamHI-NdeI
	Reverse	<u>CCCGCTCGAG</u> -CGGCGTTTATAGCGG <SEQ ID 1081>	XhoI
<b>ORF 139</b>	Forward	GCGGATCCCATATG-GCTTTTTTGGCGGTAATG <SEQ ID 1082>	BamHI-NdeI
	Reverse	<u>CCCGCTCGAG</u> -TAACGTTTCCGTGCGTTT <SEQ ID 1083>	XhoI
<b>ORF 140</b>	Forward	GCGGATCCCATATG-TTGCCACAGGCAGC <SEQ ID 1084>	BamHI-NdeI
	Reverse	<u>CCCGCTCGAG</u> -GACGATGGCAAACAGC <SEQ ID 1085>	XhoI
<b>ORF 141</b>	Forward	GCGGATCCCATATG-CCGTCTGAAGCAGTCT <SEQ ID 1086>	BamHI-NdeI
	Reverse	<u>CCCGCTCGAG</u> -ATCTGTTGTTTTTAAATATT <SEQ ID 1087>	XhoI
<b>ORF 142</b>	Forward	GCGGATCCCATATG-GATAATTCTGGTAGTGAAG <SEQ ID 1088>	BamHI-NdeI
	Reverse	<u>CCCGCTCGAG</u> -AAACGTATAGCCTACCT <SEQ ID 1089>	XhoI
<b>ORF 143</b>	Forward	GCGGATCCCATATG-GATACCGCTTTGAACCT <SEQ ID 1090>	BamHI-NdeI
	Reverse	<u>CCCGCTCGAG</u> -AATGGCTTCCGCAATATG <SEQ ID 1091>	XhoI
<b>ORF 144</b>	Forward	GCGGATCCCATATG-ACCTTTTTACAACGTTTGC <SEQ ID 1092>	BamHI-NdeI
	Reverse	<u>CCCGCTCGAG</u> -AGATTGTTGTTGTTTTTTCG <SEQ ID 1093>	XhoI
<b>ORF 147</b>	Forward	GCGGATCCCATATG-TCTGTCTTTCAAACGGC <SEQ ID 1094>	BamHI-NdeI
	Reverse	<u>CCCGCTCGAG</u> -TTTGTTTTTGCAAGACAG <SEQ ID 1095>	XhoI

NB:

- restriction sites are underlined

CHIR-0160 (356.001)

- for ORFs 110-130, where the ORF itself carries an *EcoRI* site (eg. ORF122), a *SalI* site was used in the forward primer instead. Similarly, where the ORF carries a *PstI* site (eg. ORFs 115 and 127), a *SalI* site was used in the reverse primer.

5 Oligos were synthesized by a Perkin Elmer 394 DNA/RNA Synthesizer, eluted from the columns in 2ml  $\text{NH}_4\text{OH}$ , and deprotected by 5 hours incubation at  $56^\circ\text{C}$ . The oligos were precipitated by addition of 0.3M Na-Acetate and 2 volumes ethanol. The samples were then centrifuged and the pellets resuspended in either 100 $\mu\text{l}$  or 1ml of water.  $\text{OD}_{260}$  was determined using a Perkin Elmer Lambda Bio spectrophotometer and the concentration was determined and adjusted to 2-10pmol/ $\mu\text{l}$ .

### C) Amplification

- 10 The standard PCR protocol was as follows: 50-200ng of genomic DNA were used as a template in the presence of 20-40 $\mu\text{M}$  of each oligo, 400-800 $\mu\text{M}$  dNTPs solution, 1x PCR buffer (including 1.5mM  $\text{MgCl}_2$ ), 2.5 units *TaqI* DNA polymerase (using Perkin-Elmer AmpliTaq, GIBCO Platinum, Pwo DNA polymerase, or Tahara Shuzo Taq polymerase).

In some cases, PCR was optimised by the addition of 10 $\mu\text{l}$  DMSO or 50 $\mu\text{l}$  2M betaine.

- 15 After a hot start (adding the polymerase during a preliminary 3 minute incubation of the whole mix at  $95^\circ\text{C}$ ), each sample underwent a double-step amplification: the first 5 cycles were performed using as the hybridization temperature the one of the oligos excluding the restriction enzymes tail, followed by 30 cycles performed according to the hybridization temperature of the whole length oligos. The cycles were followed by a final 10 minute extension step at  $72^\circ\text{C}$ .

- 20 The standard cycles were as follows:

	Denaturation	Hybridisation	Elongation
First 5 cycles	30 seconds $95^\circ\text{C}$	30 seconds $50-55^\circ\text{C}$	30-60 seconds $72^\circ\text{C}$
Last 30 cycles	30 seconds $95^\circ\text{C}$	30 seconds $65-70^\circ\text{C}$	30-60 seconds $72^\circ\text{C}$

The elongation time varied according to the length of the ORF to be amplified.

The amplifications were performed using either a 9600 or a 2400 Perkin Elmer GeneAmp PCR System. To check the results, 1/10 of the amplification volume was loaded onto a 1-1.5% agarose gel and the size of each amplified fragment compared with a DNA molecular weight marker.

The amplified DNA was either loaded directly on a 1% agarose gel or first precipitated with ethanol and resuspended in a suitable volume to be loaded on a 1% agarose gel. The DNA fragment corresponding to the right size band was then eluted and purified from gel, using the Qiagen Gel Extraction Kit, following the instructions of the manufacturer. The final volume of the DNA fragment was 30µl or 50µl of either water or 10mM Tris, pH 8.5.

#### D) Digestion of PCR fragments

10 The purified DNA corresponding to the amplified fragment was split into 2 aliquots and double-digested with:

- *NdeI/XhoI* or *NheI/XhoI* for cloning into pET-21b+ and further expression of the protein as a C-terminus His-tag fusion
- *BamHI/XhoI* or *EcoRI/XhoI* for cloning into pGEX-KG and further expression of the protein as N-terminus GST fusion.
- For ORF 76, *NheI/BamHI* for cloning into pTRC-HisA vector and further expression of the protein as N-terminus His-tag fusion.
- *EcoRI/PstI*, *EcoRI/SalI*, *SalI/PstI* for cloning into pGex-His and further expression of the protein as N-terminus His-tag fusion

20 Each purified DNA fragment was incubated (37°C for 3 hours to overnight) with 20 units of each restriction enzyme (New England Biolabs ) in a either 30 or 40µl final volume in the presence of the appropriate buffer. The digestion product was then purified using the QIAquick PCR purification kit, following the manufacturer's instructions, and eluted in a final volume of 30 or 50µl of either water or 10mM Tris-HCl, pH 8.5. The final DNA concentration was determined by

25 1% agarose gel electrophoresis in the presence of titrated molecular weight marker.

**E) Digestion of the cloning vectors (pET22B, pGEX-KG, pTRC-His A, and pGex-His)**

10µg plasmid was double-digested with 50 units of each restriction enzyme in 200µl reaction volume in the presence of appropriate buffer by overnight incubation at 37°C. After loading the whole digestion on a 1% agarose gel, the band corresponding to the digested vector was purified  
5 from the gel using the Qiagen QIAquick Gel Extraction Kit and the DNA was eluted in 50µl of 10mM Tris-HCl, pH 8.5. The DNA concentration was evaluated by measuring OD<sub>260</sub> of the sample, and adjusted to 50µg/µl. 1µl of plasmid was used for each cloning procedure.

The vector pGEX-His is a modified pGEX-2T vector carrying a region encoding six histidine residues upstream to the thrombin cleavage site and containing the multiple cloning site of the  
10 vector pTRC99 (Pharmacia).

**F) Cloning**

The fragments corresponding to each ORF, previously digested and purified, were ligated in both pET22b and pGEX-KG. In a final volume of 20µl, a molar ratio of 3:1 fragment/vector was ligated using  
15 0.5µl of NEB T4 DNA ligase (400 units/µl), in the presence of the buffer supplied by the manufacturer. The reaction was incubated at room temperature for 3 hours. In some experiments, ligation was performed using the Boheringer "Rapid Ligation Kit", following the manufacturer's instructions.

In order to introduce the recombinant plasmid in a suitable strain, 100µl *E. coli* DH5 competent cells were incubated with the ligase reaction solution for 40 minutes on ice, then at 37°C for 3  
20 minutes, then, after adding 800µl LB broth, again at 37°C for 20 minutes. The cells were then centrifuged at maximum speed in an Eppendorf microfuge and resuspended in approximately 200µl of the supernatant. The suspension was then plated on LB ampicillin (100mg/ml ).

The screening of the recombinant clones was performed by growing 5 randomly-chosen colonies overnight at 37°C in either 2ml (pGEX or pTC clones) or 5ml (pET clones) LB broth + 100µg/ml  
25 ampicillin. The cells were then pelleted and the DNA extracted using the Qiagen QIAprep Spin Miniprep Kit, following the manufacturer's instructions, to a final volume of 30µl. 5µl of each individual miniprep (approximately 1g ) were digested with either *NdeI/XhoI* or *BamHI/XhoI* and the whole digestion loaded onto a 1-1.5% agarose gel (depending on the expected insert size), in

parallel with the molecular weight marker (1Kb DNA Ladder, GIBCO). The screening of the positive clones was made on the base of the correct insert size.

For the cloning of ORFs 110, 111, 113, 115, 119, 122, 125 & 130, the double-digested PCR product was ligated into double-digested vector using *EcoRI-PstI* cloning sites or, for ORFs 115  
5 & 127, *EcoRI-SalI* or, for ORF 122, *SalI-PstI*. After cloning, the recombinant plasmids were introduced in the *E.coli* host W3110. Individual clones were grown overnight at 37°C in L-broth with 50µl/ml ampicillin.

### G) Expression

Each ORF cloned into the expression vector was transformed into the strain suitable for expression  
10 of the recombinant protein product. 1µl of each construct was used to transform 30µl of *E.coli* BL21 (pGEX vector), *E.coli* TOP 10 (pTRC vector) or *E.coli* BL21-DE3 (pET vector), as described above. In the case of the pGEX-His vector, the same *E.coli* strain (W3110) was used for initial cloning and expression. Single recombinant colonies were inoculated into 2ml LB+Amp (100µg/ml), incubated at 37°C overnight, then diluted 1:30 in 20ml of LB+Amp (100µg/ml) in  
15 100ml flasks, making sure that the OD<sub>600</sub> ranged between 0.1 and 0.15. The flasks were incubated at 30°C into gyratory water bath shakers until OD indicated exponential growth suitable for induction of expression (0.4-0.8 OD for pET and pTRC vectors; 0.8-1 OD for pGEX and pGEX-His vectors). For the pET, pTRC and pGEX-His vectors, the protein expression was induced by addition of 1mM IPTG, whereas in the case of pGEX system the final concentration of IPTG was  
20 0.2mM. After 3 hours incubation at 30°C, the final concentration of the sample was checked by OD. In order to check expression, 1ml of each sample was removed, centrifuged in a microfuge, the pellet resuspended in PBS, and analysed by 12% SDS-PAGE with Coomassie Blue staining. The whole sample was centrifuged at 6000g and the pellet resuspended in PBS for further use.

### H) GST-fusion proteins large-scale purification.

25 A single colony was grown overnight at 37°C on LB+Amp agar plate. The bacteria were inoculated into 20ml of LB+Amp liquid culture in a water bath shaker and grown overnight. Bacteria were diluted 1:30 into 600ml of fresh medium and allowed to grow at the optimal temperature (20-37°C) to OD<sub>550</sub> 0.8-1. Protein expression was induced with 0.2mM IPTG followed by three hours incubation. The culture was centrifuged at 8000rpm at 4°C. The supernatant was discarded and the

bacterial pellet was resuspended in 7.5ml cold PBS. The cells were disrupted by sonication on ice for 30 sec at 40W using a Branson sonifier B-15, frozen and thawed twice and centrifuged again. The supernatant was collected and mixed with 150µl Glutathione-Sepharose 4B resin (Pharmacia) (previously washed with PBS) and incubated at room temperature for 30 minutes. The sample was  
5 centrifuged at 700g for 5 minutes at 4°C. The resin was washed twice with 10ml cold PBS for 10 minutes, resuspended in 1ml cold PBS, and loaded on a disposable column. The resin was washed twice with 2ml cold PBS until the flow-through reached OD<sub>280</sub> of 0.02-0.06. The GST-fusion protein was eluted by addition of 700µl cold Glutathione elution buffer (10mM reduced glutathione, 50mM Tris-HCl) and fractions collected until the OD<sub>280</sub> was 0.1. 21µl of each fraction  
10 were loaded on a 12% SDS gel using either Biorad SDS-PAGE Molecular weight standard broad range (M1) (200, 116.25, 97.4, 66.2, 45, 31, 21.5, 14.4, 6.5 kDa) or Amersham Rainbow Marker (M2) (220, 66, 46, 30, 21.5, 14.3 kDa) as standards. As the MW of GST is 26kDa, this value must be added to the MW of each GST-fusion protein.

#### **I) His-fusion solubility analysis (ORFs 111-129)**

15 To analyse the solubility of the His-fusion expression products, pellets of 3ml cultures were resuspended in buffer M1 [500µl PBS pH 7.2]. 25µl lysozyme (10mg/ml) was added and the bacteria were incubated for 15 min at 4°C. The pellets were sonicated for 30 sec at 40W using a Branson sonifier B-15, frozen and thawed twice and then separated again into pellet and supernatant by a centrifugation step. The supernatant was collected and the pellet was resuspended  
20 in buffer M2 [8M urea, 0.5M NaCl, 20mM imidazole and 0.1M NaH<sub>2</sub>PO<sub>4</sub>] and incubated for 3 to 4 hours at 4°C. After centrifugation, the supernatant was collected and the pellet was resuspended in buffer M3 [6M guanidinium-HCl, 0.5M NaCl, 20mM imidazole and 0.1M NaH<sub>2</sub>PO<sub>4</sub>] overnight at 4°C. The supernatants from all steps were analysed by SDS-PAGE.

The proteins expressed from ORFs 113, 119 and 120 were found to be soluble in PBS, whereas  
25 ORFs 111, 122, 126 and 129 need urea and ORFs 125 and 127 need guanidinium-HCl for their solubilization.

#### **J) His-fusion large-scale purification.**

A single colony was grown overnight at 37°C on a LB + Amp agar plate. The bacteria were inoculated into 20ml of LB+Amp liquid culture and incubated overnight in a water bath shaker.

Bacteria were diluted 1:30 into 600ml fresh medium and allowed to grow at the optimal temperature (20-37°C) to OD<sub>550</sub> 0.6-0.8. Protein expression was induced by addition of 1mM IPTG and the culture further incubated for three hours. The culture was centrifuged at 8000rpm at 4°C, the supernatant was discarded and the bacterial pellet was resuspended in 7.5ml of either (i) cold  
5 buffer A (300mM NaCl, 50mM phosphate buffer, 10mM imidazole, pH 8) for soluble proteins or (ii) buffer B (urea 8M, 10mM Tris-HCl, 100mM phosphate buffer, pH 8.8) for insoluble proteins.

The cells were disrupted by sonication on ice for 30 sec at 40W using a Branson sonifier B-15, frozen and thawed two times and centrifuged again.

For insoluble proteins, the supernatant was stored at -20°C, while the pellets were resuspended in  
10 2ml buffer C (6M guanidine hydrochloride, 100mM phosphate buffer, 10mM Tris-HCl, pH 7.5) and treated in a homogenizer for 10 cycles. The product was centrifuged at 13000rpm for 40 minutes.

Supernatants were collected and mixed with 150µl Ni<sup>2+</sup>-resin (Pharmacia) (previously washed with either buffer A or buffer B, as appropriate) and incubated at room temperature with gentle agitation  
15 for 30 minutes. The sample was centrifuged at 700g for 5 minutes at 4°C. The resin was washed twice with 10ml buffer A or B for 10 minutes, resuspended in 1ml buffer A or B and loaded on a disposable column. The resin was washed at either (i) 4°C with 2ml cold buffer A or (ii) room temperature with 2ml buffer B, until the flow-through reached OD<sub>280</sub> of 0.02-0.06.

The resin was washed with either (i) 2ml cold 20mM imidazole buffer (300mM NaCl, 50mM  
20 phosphate buffer, 20mM imidazole, pH 8) or (ii) buffer D (urea 8M, 10mM Tris-HCl, 100mM phosphate buffer, pH 6.3) until the flow-through reached the O.D<sub>280</sub> of 0.02-0.06. The His-fusion protein was eluted by addition of 700µl of either (i) cold elution buffer A (300mM NaCl, 50mM phosphate buffer, 250mM imidazole, pH 8) or (ii) elution buffer B (urea 8M, 10mM Tris-HCl, 100mM phosphate buffer, pH 4.5) and fractions collected until the O.D<sub>280</sub> was 0.1. 21µl of each  
25 fraction were loaded on a 12% SDS gel.

#### **K) His-fusion proteins renaturation**

10% glycerol was added to the denatured proteins. The proteins were then diluted to 20µg/ml using dialysis buffer I (10% glycerol, 0.5M arginine, 50mM phosphate buffer, 5mM reduced glutathione, 0.5mM oxidised glutathione, 2M urea, pH 8.8) and dialysed against the same buffer at 4°C for 12-

14 hours. The protein was further dialysed against dialysis buffer II (10% glycerol, 0.5M arginine, 50mM phosphate buffer, 5mM reduced glutathione, 0.5mM oxidised glutathione, pH 8.8) for 12-14 hours at 4°C. Protein concentration was evaluated using the formula:

$$\text{Protein (mg/ml)} = (1.55 \times \text{OD}_{280}) - (0.76 \times \text{OD}_{260})$$

5 **L) His-fusion large-scale purification (ORFs 111-129)**

500ml of bacterial cultures were induced and the fusion proteins were obtained soluble in buffer M1, M2 or M3 using the procedure described above. The crude extract of the bacteria was loaded onto a Ni-NTA superflow column (Quiagen) equilibrated with buffer M1, M2 or M3 depending on the solubilization buffer of the fusion proteins. Unbound material was eluted by washing the column with the same buffer. The specific protein was eluted with the corresponding buffer containing 500mM imidazole and dialysed against the corresponding buffer without imidazole. After each run the columns were sanitized by washing with at least two column volumes of 0.5 M sodium hydroxide and reequilibrated before the next use.

**M) Mice immunisations**

20µg of each purified protein were used to immunise mice intraperitoneally. In the case of ORFs 2, 4, 15, 22, 27, 28, 37, 76, 89 and 97, Balb-C mice were immunised with Al(OH)<sub>3</sub> as adjuvant on days 1, 21 and 42, and immune response was monitored in samples taken on day 56. For ORFs 44, 106 and 132, CD1 mice were immunised using the same protocol. For ORFs 25 and 40, CD1 mice were immunised using Freund's adjuvant, rather than Al(OH)<sub>3</sub>, and the same immunisation protocol was used, except that the immune response was measured on day 42, rather than 56. Similarly, for ORFs 23, 32, 38 and 79, CD1 mice were immunised with Freund's adjuvant, but the immune response was measured on day 49.

**N) ELISA assay (sera analysis)**

The acapsulated MenB M7 strain was plated on chocolate agar plates and incubated overnight at 37°C. Bacterial colonies were collected from the agar plates using a sterile dracon swab and inoculated into 7ml of Mueller-Hinton Broth (Difco) containing 0.25% Glucose. Bacterial growth was monitored every 30 minutes by following OD<sub>620</sub>. The bacteria were let to grow until the OD reached the value of 0.3-0.4. The culture was centrifuged for 10 minutes at 10000rpm. The



supernatant was discarded and bacteria were washed once with PBS, resuspended in PBS containing 0.025% formaldehyde, and incubated for 2 hours at room temperature and then overnight at 4°C with stirring. 100µl bacterial cells were added to each well of a 96 well Greiner plate and incubated overnight at 4°C. The wells were then washed three times with PBT washing  
5 buffer (0.1% Tween-20 in PBS). 200µl of saturation buffer (2.7% Polyvinylpyrrolidone 10 in water) was added to each well and the plates incubated for 2 hours at 37°C. Wells were washed three times with PBT. 200µl of diluted sera (Dilution buffer: 1% BSA, 0.1% Tween-20, 0.1% NaN<sub>3</sub> in PBS) were added to each well and the plates incubated for 90 minutes at 37°C. Wells were washed three times with PBT. 100µl of HRP-conjugated rabbit anti-mouse (Dako) serum diluted  
10 1:2000 in dilution buffer were added to each well and the plates were incubated for 90 minutes at 37°C. Wells were washed three times with PBT buffer. 100µl of substrate buffer for HRP (25ml of citrate buffer pH5, 10mg of O-phenildiamine and 10µl of H<sub>2</sub>O) were added to each well and the plates were left at room temperature for 20 minutes. 100µl H<sub>2</sub>SO<sub>4</sub> was added to each well and OD<sub>490</sub> was followed. The ELISA was considered positive when OD<sub>490</sub> was 2.5 times the respective  
15 pre-immune sera.

#### O) FACScan bacteria Binding Assay procedure.

The acapsulated MenB M7 strain was plated on chocolate agar plates and incubated overnight at 37°C. Bacterial colonies were collected from the agar plates using a sterile dracon swab and inoculated into 4 tubes containing 8ml each Mueller-Hinton Broth (Difco) containing 0.25%  
20 glucose. Bacterial growth was monitored every 30 minutes by following OD<sub>620</sub>. The bacteria were let to grow until the OD reached the value of 0.35-0.5. The culture was centrifuged for 10 minutes at 4000rpm. The supernatant was discarded and the pellet was resuspended in blocking buffer (1% BSA, 0.4% NaN<sub>3</sub>) and centrifuged for 5 minutes at 4000rpm. Cells were resuspended in blocking buffer to reach OD<sub>620</sub> of 0.07. 100µl bacterial cells were added to each well of a Costar 96 well  
25 plate. 100µl of diluted (1:200) sera (in blocking buffer) were added to each well and plates incubated for 2 hours at 4°C. Cells were centrifuged for 5 minutes at 4000rpm, the supernatant aspirated and cells washed by addition of 200µl/well of blocking buffer in each well. 100µl of R-Phicoerytrin conjugated F(ab)<sub>2</sub> goat anti-mouse, diluted 1:100, was added to each well and plates incubated for 1 hour at 4°C. Cells were spun down by centrifugation at 4000rpm for 5 minutes and  
30 washed by addition of 200µl/well of blocking buffer. The supernatant was aspirated and cells

resuspended in 200µl/well of PBS, 0.25% formaldehyde. Samples were transferred to FACScan tubes and read. The condition for FACScan setting were: FL1 on, FL2 and FL3 off; FSC-H threshold:92; FSC PMT Voltage: E 02; SSC PMT: 474; Amp. Gains 7.1; FL-2 PMT: 539; compensation values: 0.

#### 5. **P) OMV preparations**

Bacteria were grown overnight on 5 GC plates, harvested with a loop and resuspended in 10 ml 20mM Tris-HCl. Heat inactivation was performed at 56°C for 30 minutes and the bacteria disrupted by sonication for 10 minutes on ice (50% duty cycle, 50% output). Unbroken cells were removed by centrifugation at 5000g for 10 minutes and the total cell envelope fraction recovered by centrifugation at 50000g at 4°C for 75 minutes. To extract cytoplasmic membrane proteins from the crude outer membranes, the whole fraction was resuspended in 2% sarkosyl (Sigma) and incubated at room temperature for 20 minutes. The suspension was centrifuged at 10000g for 10 minutes to remove aggregates, and the supernatant further ultracentrifuged at 50000g for 75 minutes to pellet the outer membranes. The outer membranes were resuspended in 10mM Tris-HCl, pH8 and the protein concentration measured by the Bio-Rad Protein assay, using BSA as a standard.

#### **Q) Whole Extracts preparation**

Bacteria were grown overnight on a GC plate, harvested with a loop and resuspended in 1ml of 20mM Tris-HCl. Heat inactivation was performed at 56°C for 30 minutes.

#### 20 **R) Western blotting**

Purified proteins (500ng/lane), outer membrane vesicles (5µg) and total cell extracts (25µg) derived from MenB strain 2996 were loaded on 15% SDS-PAGE and transferred to a nitrocellulose membrane. The transfer was performed for 2 hours at 150mA at 4°C, in transferring buffer (0.3 % Tris base, 1.44 % glycine, 20% methanol). The membrane was saturated by overnight incubation at 4°C in saturation buffer (10% skimmed milk, 0.1% Triton X100 in PBS). The membrane was washed twice with washing buffer (3% skimmed milk, 0.1% Triton X100 in PBS) and incubated for 2 hours at 37°C with mice sera diluted 1:200 in washing buffer. The membrane was washed twice and incubated for 90 minutes with a 1:2000 dilution of horseradish peroxidase labelled anti-

mouse Ig. The membrane was washed twice with 0.1% Triton X100 in PBS and developed with the Opti-4CN Substrate Kit (Bio-Rad). The reaction was stopped by adding water.

### S) Bactericidal assay

MC58 strain was grown overnight at 37°C on chocolate agar plates. 5-7 colonies were collected and used to inoculate 7ml Mueller-Hinton broth. The suspension was incubated at 37°C on a nutator and let to grow until OD<sub>620</sub> was 0.5-0.8. The culture was aliquoted into sterile 1.5ml Eppendorf tubes and centrifuged for 20 minutes at maximum speed in a microfuge. The pellet was washed once in Gey's buffer (Gibco) and resuspended in the same buffer to an OD<sub>620</sub> of 0.5, diluted 1:20000 in Gey's buffer and stored at 25°C.

50µl of Gey's buffer/1% BSA was added to each well of a 96-well tissue culture plate. 25µl of diluted mice sera (1:100 in Gey's buffer/0.2% BSA) were added to each well and the plate incubated at 4°C. 25µl of the previously described bacterial suspension were added to each well. 25µl of either heat-inactivated (56°C waterbath for 30 minutes) or normal baby rabbit complement were added to each well. Immediately after the addition of the baby rabbit complement, 22µl of each sample/well were plated on Mueller-Hinton agar plates (time 0). The 96-well plate was incubated for 1 hour at 37°C with rotation and then 22µl of each sample/well were plated on Mueller-Hinton agar plates (time 1). After overnight incubation the colonies corresponding to time 0 and time 1 hour were counted.

Table II gives a summary of the cloning, expression and purification results.

TABLE II – Summary of cloning, expression and purification

ORF	PCR/cloning	His-fusion expression	GST-fusion expression	Purification
orf 1	+	+	+	His-fusion
orf 2	+	+	+	GST-fusion
orf 2.1	+	n.d.	+	GST-fusion
orf 4	+	+	+	His-fusion
orf 5	+	n.d.	+	GST-fusion
orf 6	+	+	+	GST-fusion
orf 7	+	+	+	GST-fusion
orf 8	+	n.d.	n.d.	

orf 9	+	+	+	GST-fusion
orf 10	+	n.d.	n.d.	
orf 11	+	n.d.	n.d.	
orf 13	+	n.d.	+	GST-fusion
orf 15	+	+	+	GST-fusion
orf 17	+	n.d.	n.d.	
orf 18	+	n.d.	n.d.	
orf 19	+	n.d.	n.d.	
orf 20	+	n.d.	n.d.	
orf 22	+	+	+	GST-fusion
orf 23	+	+	+	His-fusion
orf 24	+	n.d.	n.d.	
orf 25	+	+	+	His-fusion
orf 26	+	n.d.	n.d.	
orf 27	+	+	+	GST-fusion
orf 28	+	+	+	GST-fusion
orf 29	+	n.d.	n.d.	
orf 32	+	+	+	His-fusion
orf 33	+	n.d.	n.d.	
orf 35	+	n.d.	n.d.	
orf 37	+	+	+	GST-fusion
orf 58	+	n.d.	n.d.	
orf 65	+	n.d.	n.d.	
orf 66	+	n.d.	n.d.	
orf 72	+	+	n.d.	His-fusion
orf 73	+	n.d.	+	n.d.
orf 75	+	n.d.	n.d.	
orf 76	+	+	n.d.	His-fusion
orf 79	+	+	n.d.	His-fusion
orf 83	+	n.d.	+	n.d.
orf 84	+	n.d.	n.d.	
orf 85	+	n.d.	+	GST-fusion
orf 89	+	n.d.	+	GST-fusion
orf 97	+	+	+	GST-fusion
orf 98	+	n.d.	n.d.	
orf 100	+	n.d.	n.d.	
orf 101	+	n.d.	n.d.	
orf 102	+	n.d.	n.d.	
orf 103	+	n.d.	n.d.	
orf 104	+	n.d.	n.d.	
orf 105	+	n.d.	n.d.	
orf 106	+	+	+	His-fusion
orf 109	+	n.d.	n.d.	
orf 110	+	n.d.	n.d.	

orf 111	+	+	n.d.	His-fusion
orf 113	+	+	n.d.	His-fusion
orf 115	n.d.	n.d.	n.d.	
orf 119	+	+	n.d.	His-fusion
orf 120	+	+	n.d.	His-fusion
orf 121	+	n.d.	n.d.	
orf 122	+	+	n.d.	His-fusion
orf 125	+	+	n.d.	His-fusion
orf 126	+	+	n.d.	His-fusion
orf 127	+	+	n.d.	His-fusion
orf 128	+	n.d.	n.d.	
orf 129	+	+	n.d.	His-fusion
orf 130	+	n.d.	n.d.	
orf 131	+	+	+	n.d.
orf 132	+	+	+	His-fusion
orf 133	+	n.d.	+	GST-fusion
orf 134	+	n.d.	n.d.	
orf 135	+	n.d.	n.d.	
orf 136	+	n.d.	n.d.	
orf 137	+	n.d.	+	GST-fusion
orf 138	+	n.d.	+	GST-fusion
orf 139	+	n.d.	n.d.	
orf 140	+	n.d.	n.d.	
orf 141	+	n.d.	n.d.	
orf 142	+	n.d.	n.d.	
orf 143	+	n.d.	n.d.	
orf 144	+	n.d.	+	n.d.
orf 147	+	n.d.	n.d.	

### Example 1

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 1>:

```

5       1  ATGAAACAGA CAGTCAA.AT GCTTGCCGCC GCCCTGATTG CCTTGGGCTT
      51  GAACCGACCG GTGTGGNCGG ATGACGTATC GGATTTTCGG GAAAACTTGC
     101  A.GCGGCAGC ACAGGGAAAT GCAGCAGCCC AATACAATT GGGCGCAATG
     151  TAT.TACAAA GGACGCGCGT GCGCCGGGAT GATGCTGAAG CGGTCAGATG
     201  GTATCGGCAG CCGGCGGAAC AGGGGTTAGC CCAAGCCCAA TACAATTTGG
     251  GCTGGATGTA TGCCAACGGG CGCGC.GTGC GCCAAGATGA TACCGAAGCG
     301  GTCAGATGGT ATCGGCAGGC GGCAGCGCAG GGGGTTGTCC AAGCCCAATA
     351  CAATTTGGGC GTGATATATG CCGAAGGACG TGGAGTGC GC CAAGACGATG
     401  TCGAAGCGGT CAGATGTTT CCGCAGGCGG CAGCGCAGGG GGTAGCCCAA
     451  GCCCAAACA ATTTGGGCGT GATGTATGCC GAAAGANCGC GCGTGCGCCA
    15   501  AGACCG...

```

This corresponds to the amino acid sequence <SEQ ID 2; ORF37>:

```

1  MKQTVXMLAA ALIALGLNRP VWXDDVSDFR ENLXAAAQGN AAAQYNLGAM
51  YXQRTVRVRD DAEAVRWYRQ PAEQGLAQAAQ YNLGWMYANG RXVRQDDTEA
101 VRWYRQAAAQ GVVQAQYNLG VIYAEGRGVR QDDVEAVRWF RQAAAQGVAAQ
151 AQNNLGVMYA ERXVRQD...

```

5 Further work revealed the complete nucleotide sequence <SEQ ID 3>:

```

1  ATGAAACAGA CAGTCAAATG GCTTGCCGCC GCCCTGATTG CCTTGGGCTT
51  GAACCGAGCG GTGTGGGCGG ATGACGTATC GGATTTTCGG GAAAACCTTGC
101 AGGCGGCAGC ACAGGGAAAT GCAGCAGCCC AATACAATTT GGGCGCAATG
151 TATTACAAAG GACGCGGCGT GCGCCGGGAT GATGCTGAAG CGGTCAGATG
10  201 GTATCGGCAG GCGGCGGAAC AGGGGTTAGC CCAAGCCCAA TACAATTTGG
251 GCTGGATGTA TGCCAACGGG CGCGGCGTGC GCCAAGATGA TACCGAAGCG
301 GTCAGATGGT ATCGGCAGGC GGCAGCGCAG GGGGTTGTCC AAGCCCAATA
351 CAATTGCGGC GTGATATATG CCGAAGGACG TGGAGTGC GC CAAGACGATG
401 TCGAAGCGGT CAGATGGTTT CGGCAGGCGG CAGCGCAGGG GGTAGCCCAA
15  451 GCGCAAAACA ATTTGGGCGT GATGTATGCC GAAAGACGCG GCGTGCGCCA
501 AGACCGCGCC CTTGCACAAG AATGGTTTGG CAAGGCTTGT CAAAACGGAG
551 ACCAAGACGG CTGCGACAAT GACCAACGCC TGAAGGCGGG TTATTGA

```

This corresponds to the amino acid sequence <SEQ ID 4; ORF37-1>:

```

1  MKQTVKWLAA ALIALGLNRA VWADDVSDFR ENLQAAAQGN AAAQYNLGAM
20  51  YYKGRGVRRD DAEAVRWYRQ AAEQGLAQAAQ YNLGWMYANG RGVRRQDDTEA
101  VRWYRQAAAQ GVVQAQYNLG VIYAEGRGVR QDDVEAVRWF RQAAAQGVAAQ
151  AQNNLGVMYA ERRGVQRDRA LAQEWFGKAC QNGDQDGDND DQRLKAGY*

```

Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 5>:

```

1  ATGAAACAGA CAGTCAAATG GCTTGCCGCC GCCCTGATTG CCTTGGGCTT
25  51  GAACCAAGCG GTGTGGGCGG ATGACGTATC GGATTTTCGG GAAAACCTTGC
101  AGGCGGCAGC ACAGGGAAAT GCAGCAGCCC AAAACAATTT GGGCGTGATG
151  TATGCCGAAA GACGCGGCGT GCGCCAAGAC CGCGCCCTTG CACAAGAATG
201  GCTTGGCAAG GCTTGTCAA ACGGATACCA AGACAGCTGC GACAATGACC
251  AACGCTGAA AGCGGGTTAT TGA

```

30 This encodes a protein having amino acid sequence <SEQ ID 6; ORF37a>:

```

1  MKQTVKWLAA ALIALGLNQA VWADDVSDFR ENLQAAAQGN AAAQNNLGVM
51  YAERRGVRRD RALAQEWL GK ACQNGYQDSC DNDQRLKAGY *

```

The originally-identified partial strain B sequence (ORF37) shows 68.0% identity over a 75aa overlap with ORF37a:

```

35  orf37.pep      10      20      30      40      50      60
               MKQTVXMLAAALIALGLNRPVWXDDVSDFR ENLXAAAQGNAAAQYNLGAMYXQRTVRVRD
               ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
40  orf37a        10      20      30      40      50      60
               MKQTVKWLAAALIALGLNQA VWADDVSDFR ENLQAAAQGNAAAQNNLGVMYAERRGVRRD
               10      20      30      40      50      60

               70      80      90      100     110     120
orf37.pep      DAEAVRWYRQPAEQGLAQAYNLGWMYANGRXVRQDDTEAVRWYRQAAAQGVVQAQYNLG
               | | : | : : |
45  orf37a        70      80      90
               RALAQEWLGKACQNGYQDSCDNDQRLKAGYX
               70      80      90

```

Further work identified the corresponding gene in *N.gonorrhoeae* <SEQ ID 7>:

```

1  ATGAAACAGA CAGTCAAATG GCTTGCCGCC GCCCTGATTG CCTTGGGCTT
50  51  GAACCAAGCG GTGTGGGCGG GTGACGTATC GGATTTTCGG GAAAACCTTGC
101  AGGCGGCAGC ACAGGGAAAT GCAGCAGCCC AATACAATTT GGGCGTGATG
151  TATGAAAATG GACAAGGAGT TCGTCAAGAT TATGTACAGG CAGTGCAGTG
201  GTATCGCAAG GCTTCAGAAC AAGGGGATGC CCAAGCCCAA TACAATTTGG
251  GCTTGATGTA TTACGATGGA CGCGGCGTGC GCCAAGACCT TGCGCTCGCT
301  CAACAATGGC TTGGCAAGGC TTGTCAAAAC GGAGACCAAA ACAGCTGCGA
55  351  CAATGACCAA CGCCTGAAGG CGGGTTATTA A

```

This encodes a protein having amino acid sequence <SEQ ID 8; ORF37ng>:

```

1  MKQTVKWLAA ALIALGLNQA VWAGDVSDFR ENLQAAEQGN AAAQFNLGVM
51 YENGQGVRRD YVQAVQWYRK ASEQGDAAQ YNLGLMYDGRGVRQDLALA
101 QQWLKGKACQN GDQNSCDNDQ RLKAGY*

```

- 5 The originally-identified partial strain B sequence (ORF37) shows 64.9% identity over a 111aa overlap with ORF37ng:

```

10 orf37.pep  MKQTVXMLAAALIALGLNRPVWXDDVSDFRNLXAAQGNAAQYNLGAMYXQRTVRVRD 60
    orf37ng    MKQTVKWLAAALIALGLNQAVWAGDVSDFRNLQAAEQGNAAQFNLGVMYENGQGVRRD 60
15 orf37.pep  DAEAVRWYRQPAEQGLAQYNLGWMYANGRXVRQDDTEAVRWYRQAAQGVVQAQYNLG 120
    orf37ng    YVQAVQWYRKASEQGDAAQYNLGMYDGRGVRQDLALAQQWLKGKACQNGDQNSCDNDQ 120
20 orf37.pep  VIYAEGRGVRQDDVEAVRWFRQAAQGVAAQNNLGVMYAERXVRQD 168
    orf37ng    RLKAGY 126

```

The complete strain B sequence (ORF37-1) and ORF37ng show 51.5% identity in 198 aa overlap:

```

20 orf37-1.pep  MKQTVKWLAAALIALGLNRAVWADDVSDFRNLQAAQGNAAQYNLGAMYKGRGVRRD 60
    orf37ng      MKQTVKWLAAALIALGLNQAVWAGDVSDFRNLQAAEQGNAAQFNLGVMYENGQGVRRD 60
25 orf37-1.pep  DAEAVRWYRQAAEQGLAQYNLGWMYANGRGVRQDDTEAVRWYRQAAQGVVQAQYNLG 120
    orf37ng      YVQAVQWYRKASEQGDAAQYNLGMYDGRGVRQD-----
30 orf37-1.pep  VIYAEGRGVRQDDVEAVRWFRQAAQGVAAQNNLGVMYAERRGVRQDRALAEWFGKAC 180
    orf37ng      -----LALAQWLKGKAC 100
35 orf37-1.pep  QNGDQDQCDNDQRLKAGYX 199
    orf37ng      QNGDQNSCDNDQRLKAGYX 120
40

```

Computer analysis of these amino acid sequences indicates a putative leader sequence, and it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

- 45 ORF37-1 (11kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 1A shows the results of affinity purification of the GST-fusion protein, and Figure 1B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result), FACS analysis (Figure 1C), and a

bactericidal assay (Figure 1D). These experiments confirm that ORF37-1 is a surface-exposed protein, and that it is a useful immunogen.

Figure 1E shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF37-1.

### Example 2

5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 9>:

10  
TTCGGCGA CATCGGCGGT TTGAAGGTCA ATGCCCCCGT CAAATCCGCA  
GGCGTATTGG TCGGGCGCGT CGGCGCTATC GGACTTGACC CGAAATCCTA  
TCAGGCGAGG GTGCGCCTCG ATTTGGACGG CAAGTATCAG TTCAGCAGCG  
ACGTTTCCCG CAAATCCTG ACTTCsGGAC TTTTGGGCGA GCAGTACATC  
GGGCTGCAGC AGGGCGGCGA CACGGAAAAC CTTGCTGCCG GCGACACCAT  
CTCCGTAACC AGTTCTGCAA TGGTTCTGGA AAACCTTATC GGCAAATTCA  
TGACGAGTTT TGCCGAGAAA AATGCCGACG GCGGCAATGC GAAAAAGCC  
GCCGAATAA

This corresponds to the amino acid sequence <SEQ ID 10>:

15  
1 FGDIGGLKVN APVKSAGVLV GRVGAIGLDP KSYQARVRD LDGKYQFSSD  
51 VSAQILTSL LGEQYIGLQQ GGDENLAAG DTISVTSSAM VLENLIGKFM  
101 TSFAEKNADG GNAEKAAE\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a hypothetical *H.influenzae* protein (ybrd.haein; accession number p45029)

20 SEQ ID 9 and ybrd.haein show 48.4% aa identity in 122 aa overlap:

25  
ybrd.h LGIGALVFLGLRVANVQGFATKSYTVTATFDNIGGLKVRAPLKIGGVVIGRVSAITLDE  
N.m FGDIGGLKVNAPVKSAGVLVGRVGAIGLDP  
ybrd.h 80 90 100 110 120 130  
KSYLPKVSIAINQEYNEIPENSSLSIKTSGLLGEQYIALTMGFDDGDTAMLKNGSQIQDT  
N.m KSYQARVRDLGKY-QFSSDVSAQILTSGLLGEQYIGLQQG---GDENLAAGDTISVT  
ybrd.h 140 150 160  
TSAMVLEDLIGQFL--YGSKKSDGNEKSESTEQ  
N.m SSAMVLENLIGKFMTSFAEKNADGGNAEKAAEX

Homology with a predicted ORF from *N.gonorrhoeae*

40 SEQ ID 9 shows 99.2% identity over a 118aa overlap with a predicted ORF from *N. gonorrhoeae*:

45  
ybrd GAAAVAFLAFRVAGGAFFGGSDKTYAVYADFGDIGGLKVNAPVKSAGVLVGRVGAIGLDP  
N.m FGDIGGLKVNAPVKSAGVLVGRVGAIGLDP  
ybrd 80 90 100 110 120 130  
KSYQARVRDLGKYQFSSDVSAQILTSGLLGEQYIGLQQGGDENLAAGDTISVTSSAM  
N.m KSYQARVRDLGKYQFSSDVSAQILTSGLLGEQYIGLQQGGDENLAAGDTISVTSSAM



yrbd  
 140 150 160  
 VLENLIGKFMTSFAEKNNAEGGNAEKAEX  
 |||||  
 5 N.m VLENLIGKFMTSFAEKNADGGNAEKAEX  
 100 110 120

The complete yrbd *H.influenzae* sequence has a leader sequence and it is expected that the full-length homologous *N.meningitidis* protein will also have one. This suggests that it is either a membrane protein, a secreted protein, or a surface protein and that the protein, or one of its epitopes, could be a useful antigen for vaccines or diagnostics.

### Example 3

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 11>:

1 ..ATTTGATAT ACCTCATCCG CAAGAATCTA GGTTCGCCCC TCTTCTTCTT  
 51 TCAGGAACGC CCCGGAAGG ACGGAAACC TTTTAAATG GTCAAATTCC  
 101 GTTCCATGCG CGACGGCTTG TATTCAGACG GCATTCCGCT GCCCGACGGA  
 151 GAACGCCTGA CACCGTTCGG CAAAAAATG CGTGCCGcCA GTwTGGACGA  
 201 ACTGCCTGAA TTATGGAATA TCTTAAAGG CGAGATGAGC CTGGTCGGCC  
 251 CCCGCCCGCT GCTGATGCAA TATCTGCCGC TGTACGACAA CTCCAAAAC  
 301 CGCCGCCACG AAATGAAACC CGGCATTACC GGCTGGGCGC AGGTCAACGG  
 351 GCGCAACGCG CTTTCGTGGG ACGAAAAAT CGCCTGCGAT GTTGGGTATA  
 401 TCGACCACTT CAGCCTGTGC CTCGACATCA AAATCCTACT GCTGACGGTT  
 451 AAAAAAGTAT TAATCAAGGA AGGGATTCC GCACAGGGCG AACA.aCCAT  
 501 GCCCCCTTTC ACAGGAAAAC GCAAACTCGC CGTCGTCCGT GCGGGCGGAC  
 551 ACGGAAAAGT CGTTGCCGAC CTTGCCGCCG CACTCGGCCG GTACAGGGAA  
 601 ATCGTTTTTC TGGACGACCG CGCACAAGGC AGCGTCAACG GCTTTTCCGT  
 651 CATCGGCACG ACGCTGCTGC TTGAAAACAG TTTATCGCCC GAACAATACG  
 701 ACGTCGCCGT CGCCGTCCGC AACAACCGCA TCCGCCGCCA AATCGCCGAA  
 751 AAAGCCGCCG CGCTCGGCTT CGCCCTGCCC GTACTGGTTC ATCCGGACGC  
 801 GACCGTCTCG CCTTTCGCAA CAGTCGGACA AGGCAGCGTC GTTATGGCGA  
 851 AAGCGTCTG..

This corresponds to the amino acid sequence <SEQ ID 12; ORF3>:

1 ..ILIYLIRKNL GSPVFFFQER PGKDGKPFKM VKFRSMRDGL YSDGIPLPDG  
 51 ERLTPFGKKL RAASXDELPE LWNILKGEMS LVGPRPLLMQ YLPLYDNFON  
 101 RRHEMKPGIT AQAQVNGRNA LSWDEKFACD VWYIDHFSLC LDIKILLTV  
 151 KKVLIKEGIS AQGEXTMPFP TGKRLAVVG AGGHGKVVD LAAALGRYRE  
 201 IVFLDDRAQG SVNGFSVIGT TLLLENSLSP EQYDVAVAVG NNRIRRQIAE  
 251 KAAALGFALP VLVHPDATVS PSATVGQGSV VMAKAV..

Further sequence analysis revealed the complete nucleotide sequence <SEQ ID 13>:

1 ATGAGTAAAT TCTTCAAACG CCTGTTTGAC ATTGTTGCCT CCGCCTCGGG  
 51 ACTGATTTTC CTCTCGCCAG TATTTTGTAT TTTGATATAC CTCATCCGCA  
 101 AGAATCTAGG TTCGCCCGTC TTCTTCTTTC AGGAACGCC CCGAAAGGAC  
 151 GGAAAACCTT TTAATATGGT CAAATTCCGT TCCATGCGCG ACGCGCTTGA  
 201 TTCAGACGGC ATTCGCTGCG CCGACGGAGA ACGCCTGACA CCGTTCGGCA  
 251 AAAAATGCG TGCGCCAGT TTGGACGAAC TGCTTGAATT ATGGAATATC  
 301 TTAATAAGCG AGATGAGCCT GGTGCGCCCC CGCCCGCTGC TGATGCAATA  
 351 TCTGCCGCTG TACGACAACT TCCAAAACCG CCGCCACGAA ATGAAACCCG  
 401 GCATTACCGG CTGGGCGCAG GTCAACGGGC GCAACGCGCT TTCGTGGGAC  
 451 GAAAAATTCG CCGCGATGT TTGGTATATC GACCACTTCA GCCTGTGCCT  
 501 CGACATCAAA ATCCTACTGC TGACGGTTAA AAAAGTATTA ATCAAGGAAG  
 551 GGATTTCCGC ACAGGGCGAA GCCACCATGC CCCCTTTCAC AGGAAAACGC  
 601 AAACCTCGCG TCGTCGGTGC GGGCGGACAC GGAAAAGTCG TTGCCGACCT  
 651 TGCCGCCGCA CTCGCCCGT ACAGGGAAT CGTTTTTCTG GACGACCGCG  
 701 CACAAGGCAG CGTCAACGGC TTTTCCGTCA TCGGCACGAC GCTGCTGCTT  
 751 GAAAACAGTT TATCGCCCGA ACAATACGAC GTCGCCGTCG CCGTCGGCAA  
 801 CAACCGCATC CGCCGCCAAA TCGCCGAAAA AGCCGCCGCG CTCGGCTTCG  
 851 CCCTGCCCGT TCTGGTTCAT CCGGACGCGA CCGTCTCGCC TTCTGCAACA

5  
 901 GTCGGACAAG GCAGCGTCGT TATGGCGAAA GCCGTCGTAC AGGCAGGCAG  
 951 CGTATTGAAA GACGGCGTGA TTGTGAACAC TGCCGCCACC GTCGATCACG  
 1001 ACTGCCTGCT TAACGCTTTC GTCCACATCA GCCCAGGCGC GCACCTGTCG  
 1051 GGCAACACGC ATATCGGCGA AGAAAGCTGG ATAGGCACGG GCGCGTGCAG  
 1101 CCGCCAGCAG ATCCGTATCG GCAGCCGCGC AACCATTGGA GCGGGCGCAG  
 1151 TCGTCGTACG CGACGTTTCA GACGGCATGA CCGTCGCGGG CAATCCGGCA  
 1201 AAGCCGCTGC CGCGCAAAAA CCCCAGAGCC TCGACAGCAT AA

This corresponds to the amino acid sequence <SEQ ID 14; ORF3-1>:

10  
 1 MSKFFKRLFD IVASASGLIF LSPVFLILY LIRKNLGSPV FFFQERPGKD  
 51 GKPFKMKVFR SMRDALDSDG IPLPDGERLT PFGKKLRAAS LDELPELWNI  
 101 LKGEMSLVGP RPLLMQYLPL YDNFQNRHE MKPGITGWAQ VNGRNALSWD  
 151 EKFCADVWYI DHFSLCLDIK ILLLTVKKVL IKEGISAQGE ATMPPTGKR  
 201 KLAVVGAGGH GKVVADLAAA LGRYREIVFL DDRAQGSVNG FSVIGTLLL  
 251 ENSLSPEQYD VAVAVGNNRI RRQIAEKAAA LGFALPVLVH PDATVSPSAT  
 15 301 VGQGSVVMK AVVQAGSVLK DGVIVNTAAT VDHDCLLNAF VHISPAHLS  
 351 GNTHIGESW IGTGACSRQQ IRIGSRATIG AGAVVVRDVS DGMTVAGNPA  
 401 KPLPRKNPET STA\*

Computer analysis of this amino acid sequence gave the following results:

# Homology with a predicted ORF from *N.meningitidis* (strain A)

20 ORF3 shows 93.0% identity over a 286aa overlap with an ORF (ORF3a) from strain A of *N. meningitidis*:

25  
 orf3.pep ILLIYLIRKNLGSPVFFFQERPGKDGKPFKMKVFR  
 orf3a MSKFFKRLFDIVASASGLIFLSPVFLILY LIRKNLGSPVFFFQERPGKDGKPFKMKVFR  
 30  
 orf3.pep SMRDGLYSDGIPLPDGERLTPFGKKLRAASXDELPELWNILKGEMSLVGP RPLLMQYLPL  
 orf3a SMHDALDSDGILLPDGERLTPFGKKLRAASLDELPELWNVLKGDMSLVGP RPLLMQYLPL  
 35  
 orf3.pep YDNFQNRHEMKPGITGWAQVNGRNALSWDEKFCADVWYIDHFSLCLDIKILLTVKKVL  
 orf3a YDNFQNRHEMKPGITGWAQVNGRNALSWDERFACDIWYIDHFSLCLDIKILLTVKKVL  
 40  
 orf3.pep IKEGISAQGEATMPPTGKRKLAVVGAGGHGKVVADLAAALGRYREIVFLDDRAQGSVNG  
 orf3a IKEGISAQGEATMPPTGKRKLAVVGAGGHGKVVADLAAALGRYREIVFLDDRAQGSVNG  
 45  
 orf3.pep FSVIGTLLLENSLSPEQYDVAVAVGNNRI RRQIAEKAAA LGFALPVLVHPDATVSPSAT  
 orf3a FSVIGTLLLENSLSPEQYDVAVAVGNNRI RRQIAEKAAA LGFALPVLVHPDATVSPSAT  
 50  
 orf3.pep VGQGSVVMKAV  
 orf3a VGQGSVVMKAVVQADSVLKDGVIVNTAATVDHDCLLDAFVHISPAHLSGNTRIGESW  
 55  
 orf3a VGQGSVVMKAVVQADSVLKDGVIVNTAATVDHDCLLDAFVHISPAHLSGNTRIGESW

The complete length ORF3a nucleotide sequence <SEQ ID 15> is:

60  
 1 ATGAGTAAAT TCTTCAAACG CCTGTTTGAC ATTGTTGCCT CCGCCTCGGG  
 51 ACTGATTTTC CTCTCGCCAG TATTTTGTAT TTTGATATAC CTCATCCGCA  
 101 AGAATCTGGG TTCGCCCGTC TTCTTCTTTC AGGAACGCC CCGAAAGGAC

5  
10  
15  
20

```

151  GGAAACCTT  TTAAATGGT  CAAATTCCGT  TCCATGCACG  ACGCGCTTGA
201  TTCAGACGGC  ATTCTGCTGC  CCGACGGAGA  ACGCCTGACA  CCGTTCGGCA
251  AAAAAGTGGC  TGCCGCCAGT  TTGGACGAAC  TGCCCGAACT  GTGGAACGTC
301  CTCAAAGGCG  ACATGAGCCT  GGTGCGCCCC  CGCCCGCTGC  TGATGCAATA
351  TCTGCGCTG  TACGACAACT  TCCAAACCG  CCGCCACGAA  ATGAAACCGG
401  GCATTACCGG  CTGGGCGCAG  GTCAACGGGC  GCAACGCGCT  TTCTGGGAC
451  GAACGCTTCG  CATGCGACAT  CTGGTATATC  GACCACTTCA  GCCTGTGCCT
501  CGACATCAAA  ATCCTACTGC  TGACGGTTAA  AAAAGTATTA  ATCAAAGAAG
551  GGATTTCGCG  ACAGGCGGAA  GCCACCATGC  CCCCTTTCAC  AGGAAAACGC
601  AAACCTGCGG  TCGTCGGTGC  GGGCGGACAC  GGCAAAGTCG  TTGCCGAGCT
651  TGCCGCGGCA  CTCGGCACAT  ACGGCGAAAT  CGTTTTTCTG  GACGACCGCG
701  TCCAAGGCAG  CGTCAACGGC  TTCCCGCTCA  TCGGCACGAC  GCTGCTGCTT
751  GAAAACAGTT  TATCGCCCGA  ACAATTCGAC  ATCGCCGTCG  CCGTCGGCAA
801  CAACCGCATC  CGCCGCCAAA  TCGCCGAAAA  AGCCGCCGCG  CTCGGCTTCG
851  CCCTGCCCGT  CCTGATTTCAT  CCGGACTCGA  CCGTCTCGCC  TTCTGCAACA
901  GTCGGACAAG  GCGGCGTCGT  TATGGCGAAA  GCCGTCGTAC  AGGCTGACAG
951  CGTATTGAAA  GACGGCGTAA  TTGTGAACAC  TGCCGCCACC  GTCGATCACG
1001 ATTGCCTGCT  TGATGCTTTC  GTCCACATCA  GCGCGGCGCG  GCACCTGTCTG
1051 GGCAACACGC  GTATCGGCGA  AGAAAGCTGG  ATAGGCACAG  GCGCGTGCAG
1101 CCGCCAGCAG  ATCCGTATCG  GCAGCCGCGC  AACCATTGGA  GCGGGCGCAG
1151 TCGTCGTGCG  CGACGTTTCA  GACGGCATGA  CCGTCGCGGG  CAACCCGGCA
1201 AAACCATTGG  CAGGCAAAAA  TACCGAGACC  CTGCGGTCGT  AA

```

This is predicted to encode a protein having amino acid sequence <SEQ ID 16>:

25  
30

```

1  MSKFFKRLFD  IVASASGLIF  LSPVFLILY  LIRKNLGSVP  FFFQERPGKD
51  GKPFKMKVFR  SMHDALSDG  ILLPDGERLT  PFGKKLRAAS  LDELPELWNV
101 LKGDMSLVGP  RPLLMQYLPL  YDNFQNRHE  MKPGITGWAQ  VNGRNALSWD
151 ERFACDIWYI  DHFSLCLDIK  ILLTVKKVL  IKEGISAQGE  ATMPPTGKR
201 KLAVVGAGGH  GKVVAAELAA  LGTYGEIVFL  DDRVQGSVNG  FPIVIGTLLL
251 ENSLSPEQFD  IAVAVGNRI  RRQIAEKAAA  LGFALPVLIH  PDSTVSPSAT
301 VGQGGVMAK  AVQADSVLK  DGIVNTAAT  VDHDCLLDAF  VHISFGAHL
351 GNTRIGESW  IGTGACSRQQ  IRIGSRATIG  AGAVVVRDVS  DGMTVAGNPA
401 KPLAGKNTET  LRS*

```

Two transmembrane domains are underlined.

ORF3-1 shows 94.6% identity in 410 aa overlap with ORF3a:

35  
40  
45  
50  
55  
60

```

      10      20      30      40      50      60
orf3a.pep  MSKFFKRLFDIVASASGLIFLSPVFLILYIIRKNLGSVPFFFQERPGKDGKPFKMKVFR
orf3-1      MSKFFKRLFDIVASASGLIFLSPVFLILYIIRKNLGSVPFFFQERPGKDGKPFKMKVFR
      10      20      30      40      50      60
orf3a.pep  SMHDALSDGILLPDGERLTPFGKKLRAASLDELPELWNVLKGDMSLVGPRPLLMQYLPL
orf3-1      SMRDALSDGILPLDGERLTPFGKKLRAASLDELPELWNILKGEMSLVGPRPLLMQYLPL
      70      80      90      100     110     120
orf3a.pep  YDNFQNRHEMKPGITGWAQVNGRNALSWEKDFACDIWYIDHFSCLDIKILLTVKKVL
orf3-1      YDNFQNRHEMKPGITGWAQVNGRNALSWEKDFACDVWYIDHFSCLDIKILLTVKKVL
      130     140     150     160     170     180
orf3a.pep  IKEGISAQGEATMPPTGKRKLAVVGAGGHGKVVAAELAAALGTYGEIVFLDDRQGSVNG
orf3-1      IKEGISAQGEATMPPTGKRKLAVVGAGGHGKVVADLAAALGRYREIVFLDDRQGSVNG
      190     200     210     220     230     240
orf3a.pep  FPIVIGTLLLENSLSPEQFDIAVAVGNRIIRQIAEKAAAALGFALPVLVHPDSTVSPSAT
orf3-1      FSVIGTLLLENSLSPEQYDVAVAVGNRIIRQIAEKAAAALGFALPVLVHPDSTVSPSAT
      250     260     270     280     290     300

```

		310	320	330	340	350	360
	orf3a.pep	VGQGGVVM	AKAVVQADS	VLKDG	VIVNTAAT	VDHDC	LLDAFVHIS
5	orf3-1	VGQGGVVM	AKAVVQADS	VLKDG	VIVNTAAT	VDHDC	LLDAFVHIS
		310	320	330	340	350	360
	orf3a.pep	IGTGACSR	QQIRIGSR	ATIGAG	AVVVRD	VS	SDGMTVAG
10	orf3-1	IGTGACSR	QQIRIGSR	ATIGAG	AVVVRD	VS	SDGMTVAG
		370	380	390	400	410	

### Homology with hypothetical protein encoded by *yvfc* gene (accession Z71928) of *B. subtilis*

15 ORF3 and YVFC proteins show 55% aa identity in 170 aa overlap (BLASTp):

	ORF3	3	IYLIRK	NLGSPV	FFFQER	PGKDG	KPFKM	VKFRSM	RDGLY	SDGIPL	PDGERL	TPFGK	KLRA	62
			I ++R	+GSPVEF	Q	RPG	GKPF	+ KFR+M	D	S	G	LPD	RLT	G+ +R
	yvfc	27	IYVRLK	IGSPVF	FKQVRP	GLHGK	PFTLY	KFRMT	TDERDS	KGNLLP	DEVRLT	KTGR	LIRK	86
20	ORF3	63	ASXDEL	PELWN	ILKGEM	SLVGPR	PLLMQ	YLYPL	YDNFQ	NRHMK	PGITG	WAQVN	GRNAL	122
			S	DELP+L	N+LKG++	SLVGPR	PLLM	YLYPL	Q	RRHE+K	PGITG	WAQ+N	GRNA+S	
	yvfc	87	LSIDEL	PQLLN	NVLKGD	LSLVGP	RPLMD	YLYPL	YTEQA	RRHEV	KPGITG	WAQIN	GRNAIS	146
25	ORF3	123	WDEKFA	CDVWY	IDHFS	LC	LDXXXXXXXXXXXXXXXXX	EGISA	QGE	XTMP	PFTG	172		
			W++KF	DVWY+D++S	LD			EGI		T	FTG			
	yvfc	147	WEKKFE	LDVWY	VDNWS	FFLDL	KILCL	TVRKV	LVSE	GIQQT	NHVT	AERFTG	196	

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF3 shows 86.3% identity over a 286aa overlap with a predicted ORF (ORF3.ng) from *N. gonorrhoeae*:

	orf3		ILIIYLIRK	NLGSPV	FFFQER	PGKDG	KPFKM	VKFR	34
	orf3ng		MSKAVK	RLFDII	ASAGL	IVLSP	VLVIYL	IRKNGS	60
35	orf3		SMRDGL	YSDGI	PLPDGE	RLTPF	GKKLRA	ASXDEL	94
	orf3ng		SMRDAL	DSDGI	PLPDSER	LTDFG	KKLRAT	SLDEL	120
40	orf3		YDNFQ	NRHMK	PGITG	WAQVN	GRNAL	SWDEK	154
	orf3ng		YNKFQ	NRHMK	PGITG	WAQVN	GRNAL	SWDEK	180
45	orf3		IKEGISA	QGE	XTMP	PFTGKR	KLAVG	AGGHG	214
	orf3ng		IKEGISA	QGE	ATMP	PFAGN	RKLAV	GAGGH	240
50	orf3		FSVIGT	TLL	ENSL	SPEQY	DVAV	AVGN	274
	orf3ng		FPVIGT	TLL	ENSL	SPEQD	ITVAV	GNRIR	300
	orf3		VGQGS	VMAK	AV				286
	orf3ng		IGQGS	VMAK	AVVQ	AGSVL	KDGV	IVNTA	360

The complete length ORF3ng nucleotide sequence <SEQ ID 17> is:

55	1	ATGAGTAAAG	CCGTCAAACG	CCTGTTTCGAC	ATCATCGCAT	CCGCATCGGG
	51	GCTGATTGTC	CTGTCGCCCG	TGTTTTTGGT	TTAATATAC	CTCATCCGCA
	101	AAACTTAGG	TTCGCCCGTC	TTCTTCattC	GGGAACGCCc	cgGAAAGGAc
	151	ggaaaacCTT	TTAAATGGT	CAAATTCCTG	TCCAtgcgcg	acgcgcttGA
60	201	TTCAGACGGC	ATTCCGCTGC	CCGATAGCGA	ACGCCTGACC	GATTTCGGCA
	251	AAAAATTACG	CGCCACCACT	TTGGACGAAC	TTCCTGAATT	ATGGAATGTC
	301	CTCAAAGCG	AGATGAGCCT	GGTCGGCCCC	CGCCCGCTT	TGATGCAGTA

5  
10  
15

```

351 TCTGCCGCTT TACAACAAAT TTCAAAACCG CCGCCACGAA ATGAAACCGG
401 GCATTACCGG CTGGGCGCAG GTCAACGGGC GCAACGCGCT TTCGTGGGAC
451 GAAAAGTTCT CCTGCGATGT TTGGTACACC GACAATTCA GCTTTTGGCT
501 GGATATGAAA ATCCTGTTTC TGACAGTCAA AAAAGTCTTG ATTAAGAAG
551 GCATTTCGGC GCAAGGGGAA GCCACCATGC CCCCTTTCGC GGGGAATCGC
601 AAATCGCCG TTATCGGCGC GGGCGGACAC GGCAAAGTCG TTGCCGAGCT
651 TGCCGCGCA CTCGCGACAT ACGGCGAAAT CGTTTTCTG GACGACCGCA
701 CCCAAGGCAG CGTCAACGGC TTCCCCGTCA TCGGCACGAC GCTGCTGCTT
751 GAAAACAGTT TATCGCCCGA ACAATTCGAC ATCACCCTCG CCGTCGGCAA
801 CAACCGCATC CGCCGCCAAA TCACCGAAAA CGCCGCCGCG CTCGGCTTCA
851 AACTGCCCGT TCTGATTCAT CCCGACGCGA CCGTCTCGCC TTCTGCAATA
901 ATCGGACAAG GCAGCGTCGT AATGGCGAAA GCCGTCGTAC AGGCCGGCAG
951 CGTATTGAAA GACGCGTGA TTGTGAACAC TGCCGCCACC GTCGATCAGC
1001 ACTGCCTGCT TGACGCTTTC GtccaCATCA GCCCGGGCGC GCACCTGTCG
1051 GGCAACACGC GTATCGGCGA AGAAAGCCGG ATAGGCACGG GCGCGTGCAG
1101 CCGCCAGCAG ACAACCGTCG GCAGCGGGGT TACCgccgGT GCAGGGgCGG
1151 TTATCGTATG CGACATCCCG GACGGCATGA CCGTCGCGGG CAACCCGGCA
1201 AAGCCCCTTA CGGGCAAAAA CCCCAGACC GGGACGGCAT AA

```

This encodes a protein having amino acid sequence <SEQ ID 18>:

20  
25

```

1 MSKAVKRLFD IIASASGLIV LSPVFLVIY LIRKNLGSPV FFIRERPGKD
51 GKPFKMVKFR SMRDALDSDG IPLPDSERLT DFGKKLRATS LDELPELWNV
101 LKGEMSLVGP RPLLMQYLPL YNKFQNRHE MKPGITGWAQ VNGRNALSWD
151 EKFSQDVWYT DNFSFWLDMK ILFLTVMKVL IKEGISAQGE ATMPFFAGNR
201 KLAVIGAGGH GKVVAELAAA LGTYGEIVFL DDRTQGSVNG FVIGTTLTL
251 ENSLSPEQFD ITVAVGNRI RRQITENAAA LGFKLPVLIH PDATVSPSAI
301 IGQGSVVMK AVVQAGSVLK DGVIVNTAAT VDHDCLDAF VHISPGAHLS
351 GNTRIGESR IGTGACSRQ TTGSGVTAG AGAVIVCDIP DGMTVAGNPA
401 KPLTGKNPKT GTA*

```

This protein shows 86.9% identity in 413 aa overlap with ORF3-1:

30  
35  
40  
45  
50  
55  
60  
65

```

      10      20      30      40      50      60
orf3-1.pep MSKFFKRLFDIVASASGLIFLSPVFLIYLIYLRKNLGSPVFFFQERPGKDGGKPFKMVKFR
      ||| |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||
orf3ng MSKAVKRLFDIIASASGLIVLSPVFLVIYLRKNLGSPVFFIRERPGKDGGKPFKMVKFR
      10      20      30      40      50      60

      70      80      90      100     110     120
orf3-1.pep SMRDALDSDGIPLPDSERLTPFGKKLRAASLDELPELWNILKGEMSLVGPRLLMQYLPL
      ||||| |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||
orf3ng SMRDALDSDGIPLPDSERLTDGKLRATSDELPELWNVLKGEMSLVGPRLLMQYLPL
      70      80      90      100     110     120

      130     140     150     160     170     180
orf3-1.pep YDNFQNRHEMKPGITGWAQVNGRNALSWDEKFACDVWYIDHFSCLDIKILLTVKKVL
      |:| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf3ng YNKFQNRHEMKPGITGWAQVNGRNALSWDEKFSCDVWYTDNFSFWLDMKILFLTVMKVL
      130     140     150     160     170     180

      190     200     210     220     230     240
orf3-1.pep IKEGISAQGEATMPFFTGRKLA VVGAGGHGKVVA DLAAALGRYREIVFLDDRAQGSVNG
      ||||| |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||
orf3ng IKEGISAQGEATMPFFAGNRKLAVIGAGGHGKVVAELAAALGTYGEIVFLDDRTQGSVNG
      190     200     210     220     230     240

      250     260     270     280     290     300
orf3-1.pep FSVIGTTLLENSLSPEQYDVAVAVGNRIIRQIAEKAAALGFALPVLVHPDATVSPSAT
      | ||||| |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||
orf3ng FSVIGTTLLENSLSPEQFDITVAVGNRIIRQITENAAALGFALPVLVHPDATVSPSAI
      250     260     270     280     290     300

      310     320     330     340     350     360
orf3-1.pep VGQGSVVMKAVVQAGSVLKDGVIVNTAATVDHDCLLNAFVHISPGAHLSGNTHIGESW
      : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf3ng IGQGSVVMKAVVQAGSVLKDGVIVNTAATVDHDCLLDAFVHISPGAHLSGNTRIGESR
      310     320     330     340     350     360

      370     380     390     400     410

```

```

orf3-1.pep      IGTGACSRQQIRIGSRATIGAGAVVVRDVS DGMTVAGNPAKPLPRKNPETSTAX
                  |||||  |||  :||  :|  |||||  :|  :|||  |||||  |||  :|||
orf3ng          IGTGACSRQQTTVGSGVTAGAGAVIVCDIPDGMTVAGNPAKPLTGKNPKTGTX
                  370      380      390      400      410

```

5 In addition, ORF3ng shows significant homology with a hypothetical protein from *B.subtilis*:

```

gnl|PID|e238668 (Z71928) hypothetical protein [Bacillus subtilis]
>gi|1945702|gnl|PID|e313004 (Z94043) hypothetical protein [Bacillus subtilis]
>gi|2635938|gnl|PID|e186113 (Z99121) similar to capsular polysaccharide
biosynthesis [Bacillus subtilis]Length = 202
Score = 235 bits (594), Expect = 3e-61
Identities = 114/195 (58%), Positives = 142/195 (72%)

```

15	Query: 5	VKRLFDIIASASGLIVLSPVFLVLIYLIRKNLGSVPVFFIRERPGKGKPFKMKVFRSMRD	64
	Sbjct: 3	+KRLFD+ A+ L S + L I ++R +GSPVFF + RPG GKPF + KFR+M D	62
20	Query: 65	ALDSDGIPLPDSERLTDGFKKLRSLSDELPELWNVLKGEMSLVGPRPLLMOYLPLYNKF	124
	Sbjct: 63	DS G LPD RLT G+ +R S+DELP+L NVLKG++SLVGPRPLLMOYLPLYNKF	122
25	Query: 125	QNRHRHEMKPGITGWAQVNGRNASLWDEKFSCDVWYTDNFSFWLDMKILFLTVMKVLKEG	184
	Sbjct: 123	QARRHEVKGPGITGWAQINGRNASLWDEKFSCDVWYVDNWSFFLDLKLCLTVRKVLVSEG	182

The hypothetical product of *yvfC* gene shows similarity to EXOY of *R.meliloti*, an  
30 exopolysaccharide production protein. Based on this and on the two predicted transmembrane  
regions in the homologous *N.gonorrhoeae* sequence, it is predicted that these proteins, or their  
epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 4

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 19>:

35                    1    ..AACCATATGG CGATTGTCAT CGACGAATAC GGC GG CACAT CCGGCTTGGT  
                     51    CACCTTTGAA GACATCATCG AGCAAATCGT CGGCGAAATC GAAGACGAGT  
                    101    TTGACGAAGA CGATAGCGCC GACAATATCC ATGCCGTTC TTCAGACACG  
                    151    TGGCGCATCC ATGCAGCTAC CGAAATCGAA GACATCAACA CCTTCTTCGG  
                    201    CACGCGAATC AGCATCGAAG AAGCCGACAC CATT.GGCGG CTGGTTCATT  
40                   251    CAAGAGTTGG GACATCTGCC CGTGCGCGGC GAAAAAGTCC TTATCGGCGG  
                    301    TTTGCAGTTC ACCGTCGCAC GCGCCGACAA CCGCCGCCTG CATACGCTGA  
                    351    TGGCGACCCG CGTGAAGTAA GC..... ACCGC CGTTTCTGCA  
                    401    CAGTTTAG

This corresponds to amino acid sequence <SEQ ID 20; ORF5>:

```

45      1  ..NHMAIVIDEY GGTSGLVTFE DIIEQIVGEI EDEFDEDDSA DNIHAVSSDT
      51  WRIHAATEIE DINTFFGTEY SIEEADTIXR PGHSRVGTSA RARRKSPYRR
     101  FAVHRRTRRQ PPPAYADGDP REVS...XR RFCTV*

```

Further sequence analysis revealed the complete DNA sequence to be <SEQ ID 21>:

50

1	ATGGACGGCG	CACAACCGAA	AACGAATTTT	TTTGAACGCC	TGATTGCCCG
51	ACTCGCCCGC	GAACCCGATT	CCGCCGAAGA	CGTATTAAAC	CTGCTTCGGC
101	AGGGCGACGA	CGAGGAAGTT	TTTGATGCGG	ATACGCTTTT	AAGATTGGAA
151	AAAGTCTCTG	ATTTTTCCTG	TTTGAAGTGC	CGCGACGCGA	TGATTACGCG
201	CAGCCGTATG	AACGTTTAA	AAGAAAACGA	CAGCATCGAG	CGCATCAACG
251	CCTACGTTAT	CGATACCGCC	CATTGCGGCT	TCCCGGTCAT	CGCGCAAGAC

5  
10  
301 AAAGACGAAG TTTTGGGCAT TTTGCACGCC AAAGACCTGC TCAAATATAT  
351 GTTTAACCCC GAGCAGTTCC ACCTCAAATC CATTCTCCGC CCCGCCGTCT  
401 TCGTCCCCGA AGGCAAATCG CTGACCGCCC TTTTAAAAGA GTTCCGCGAA  
451 CAGCGCAACC ATATGGCGAT TGTATCGAC GAATACGGCG GCACATCCGG  
501 CTTGGTCACC TTTGAAGACA TCATCGAGCA AATCGTCGGC GAAATCGAAG  
551 ACAGTTTGA CGAAGACGAT AGCGCCGACA ATATCCATGC CGTTTCTTCC  
601 GAACGCTGGC GCATCCATGC AGCTACCGAA ATCGAAGACA TCAACACCTT  
651 CTTCCGGCAG GAATACAGCA GCGAAGAAGC CGACACCATT CGGCCTGGTC  
701 ATTCAAGAGT TGGGACATCT GCCCGTGC GC CGGAAAAAG TCCTTATCGG  
751 CGGTTTGCAG TTCACCGTCG CACGCGCCGA CAACCGCCGC CTGCATACGC  
801 TGATGGCGAC CCGCGTGAAG TAAGCACCGC CGTTTCTGCA CAGTTTAGGA  
851 TGACGGTACG GCGTTTTCT GTTTCAATCC GCCCCATCCG CCAAACATAA

This corresponds to amino acid sequence <SEQ ID 22; ORF5-1>:

15  
1 MDGAQPKTNF FERLIARLAR EPDSAEDVLN LLRQAHEQEV FDADTLLRLE  
51 KVLDFSDLEV RDAMITRSRM NVLKENDSIE RITAYVIDTA HSRFPVIGED  
101 KDEVLGILHA KDLLKYMFP EQFHLKSILR PAVFVPEGKS LTALLKEFRE  
151 QRNHMAIVID EYGGTSGLVT FEDIIIEQIVG EIEDEFDEDD SADNIHAVSS  
201 ERWRIHAATE IEDINTFFGT EYSSEADTI RPHGSRVGT ARARRKSPYR  
251 RFAVHRRTRR QPPPAYADGD PREVSTAVSA QFRMTVRAF VSIRPIRQT\*

20 Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 23 >:

25  
30  
35  
1 ATGGACGGCG CACAACCGAA AACAAATTT TTNNAACGCC TGATTGCCCG  
51 ACTCGCCCGC GAACCCGATT CCGCCGAAGA CGTATTGACC CTGTTGCGCC  
101 AAGCGCACGA ACAGGAAGTA TTTGATGCGG ATACGCTTTT AAGATTGGAA  
151 AAAGTCCTCG ATTTTCTCTGA TTTGGAAGTG CGCGACGCGA TGATTACGCG  
201 CAGCCGTATG AACGTTTAA AAGAAACGA CAGCATCGAA CGCATCACCG  
251 CCTACGTTAT CGATACCGCC CATTGCGCT TCCCCGTAT CGGTGAAGAC  
301 AAAGACGAAG TTTTGGGTAT TTTGCACGCC AAAGACCTGC TCAAATATAT  
351 GTTCAACCCC GAGCAGTTCC ACCTCAAATC GATATTGCGC CCTGCCGTCT  
401 TCGTCCCCGA AGGCAAATCG CTGACCGCCC TTTTAAAAGA GTTCCGCGAA  
451 CAGCGCAACC ATATGGCAAT CGTCATCGAC GAATACGCGC GCACGTCGGG  
501 TTTGGTAACT TTTGAAGACA TCATCGAGCA AATCGTCGGC GACATCGAAG  
551 ATGAGTTTGA CGAAGACGAA AGCGCGGACA ACATCCACGC CGTTTCCGCC  
601 GAACGCTGGC GCATCCACGC GGCTACCGAA ATCGAAGACA TCAACGCCTT  
651 TTTCCGGCACG GAATACAGCA GCGAAGAAGC CGACACCATC GGCGGCCNTG  
701 GTCATTGAGG AATTGGNACA CCTGCCCGTG CGCGGCGAAA AAGTCNTTAT  
751 CGGCGNNTTG CANTTACNG TCGCCNGCGC NGACAACCGC CGCCTGCATA  
801 CGCTGATGGC GACCCGCGTG AAGTAAGCTC CGCCGTTTCT GTACAGTTTA  
851 GGATGACGGT ACGGGCGTTT TCTGTTTCAA TCCGCCCAT CCGCCANACA  
901 TAA

40 This encodes a protein having amino acid sequence <SEQ ID 24; ORF5a>:

45  
1 MDGAQPKTNF XXRLIARLAR EPDSAEDVLT LLRQAHEQEV FDADTLLRLE  
51 KVLDFSDLEV RDAMITRSRM NVLKENDSIE RITAYVIDTA HSRFPVIGED  
101 KDEVLGILHA KDLLKYMFP EQFHLKSILR PAVFVPEGKS LTALLKEFRE  
151 QRNHMAIVID EYGGTSGLVT FEDIIIEQIVG DIEDEFDEDE SADNIHAVSA  
201 ERWRIHAATE IEDINAFPGT EYSSEADTI GGXGHSIGT PARARRKSYX  
251 RRXAXHXR XR QPPPAYADG DPREVSSAVS VQFRMTVRAF SVSIRPIRXT  
301 \*

The originally-identified partial strain B sequence (ORF5) shows 54.7% identity over a 124aa overlap with ORF5a:

50  
orf5.pep  
orf5a  
55  
orf5.pep  
orf5a  
60

				10	20	30
				NHMAIVIDEYGGTSGLVT	FEDIIIEQIVGEI	
	FHLKSILRPAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGLVT	FEDIIIEQIVGDI				
	130	140	150	160	170	180
		40	50	60	70	80
	EDEFDEDD	SADNIHAVSS	DTWRIHAATE	IEDINTFFGT	EYSIEEADTI	KRPHGSRVGTSA
	EDEFDEDESADNIHAVSAERWRIHAATE	IEDINAFPGT	EYSSEADTI	GGXGHSIGT	PA	
	190	200	210	220	230	240

orf5.pep RARRKSPYRRFAVHRRTRRQPPPAYADGDPREVSSXXXRRFCTV  
 100 110 120 130  
 5 orf5a RARRKSXYRRXAXHXRXXRQPPPAYADGDPREVSSAVSVQFRMTVRAFSVSIRPIRXTX  
 250 260 270 280 290 300

The complete strain B sequence (ORF5-1) and ORF5a show 92.7% identity in 300 aa overlap:

10 orf5a.pep MDGAQPKTNFXXRLIARLAREPDSAEDVLTLLRQAHEQEVFDADTLLRLEKVLDFSDLEV  
 orf5-1 MDGAQPKTNFFERLIARLAREPDSAEDVLNLLRQAHEQEVFDADTLLRLEKVLDFSDLEV  
 15 orf5a.pep RDAMITRSRMNVLKENDSIERITAYVIDTAHSRFPVIGEDKDEVLGILHAKDLLKYMFP  
 orf5-1 RDAMITRSRMNVLKENDSIERITAYVIDTAHSRFPVIGEDKDEVLGILHAKDLLKYMFP  
 20 orf5a.pep EQFHLKSILRPVAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGGLVTFEDIIEQIVG  
 orf5-1 EQFHLKSILRPVAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGGLVTFEDIIEQIVG  
 25 orf5a.pep DIEDEFDEDESADNIHAVSAERWRIHAATEIEDINAFGTEYSSEEADTIGGXGHSIGT  
 orf5-1 EIEDEFDEDDSDADNIHAVSSERWRIHAATEIEDINTFFGTEYSSEEADTIRP-GHSRVGT  
 30 orf5a.pep PARARRKSXYRRXAXHXRXXRQPPPAYADGDPREVSSAVSVQFRMTVRAFSVSIRPIRXT  
 orf5-1 SARARRKSXYRRFAVHRRTRRQPPPAYADGDPREVSTAVSAQFRMTVRAFSVSIRPIRQT  
 240 250 260 270 280 290

Further work identified the a partial DNA sequence in *N.gonorrhoeae* <SEQ ID 25> which encodes a protein having amino acid sequence <SEQ ID 26; ORF5ng>:

1 MDGAQPKTNF FERLIARLAR EPDSAEDVLN LLRQAHEQEV FDADTLTRLR  
 40 51 KVLDFAELEV RDAMITRSRM NVLKENDSIE RITAYVIDTA HSRFPVIGED  
 101 KDEVLGILHA KDLLKYMFP EQFHLKSVLR PAVFVPEGKS LTALLKEFRE  
 151 QRNHMAIVID EYGGTSGGLV FEDIIEQIVG DIEDEFDEDE SADDIHSVSA  
 201 ERWRIHAATE IEDINAFGT EYGSEEADTI RRLGHSIGT PARARRKSPY  
 251 RRFAVHRRPR RQPPPAHADG DPREVSRACP HRRFCTV\*

Further analysis revealed the complete gonococcal nucleotide sequence <SEQ ID 27> to be:

1 ATGGACGGCG CACAACCGAA AACAAATTTT TTTGAACGCC TGATTGCCCG  
 51 ACTCGCCCGC GAACCCGATT CCGCCGAAGA CGTATTAAAC CTGCTTCGGC  
 101 AGGCGCACGA ACAGGAAGTT TTTGATGCCG ACACACTGAC CCGGCTGGAA  
 151 AAAGTATTGG ACTTTGCCGA GCTGGAAGTG CGCGATGCCA TGATTACGGC  
 50 201 CAGCCGCATG AACGTATTGA AAGAAAACGA CAGCATCGAA CGCATCACCG  
 251 CCTACGTCAT CGATACCGCC CATTCGCGCT TCCCCGTCAT CGGCGAAGAC  
 301 AAAGACGAAG TTTTGGGCAT TTTGCACGCC AAAGACCTGC TCAAATATAT  
 351 GTTCAACCCC GAGCAGTTCC ACCTGAAATC CGTCTTGCCG CCTGCCGTTT  
 401 TCGTGCCCGA AGGCAAATCT TTGACCGCCC TTTTAAAGA GTTCCGCGAA  
 451 CAGCGCAACC ATATGGCAAT CGTCATCGAC GAATACGGCG GCACGTCGGG  
 55 501 TTTGGTCACC TTGAGAAGCA TCATCGAGCA AATCGTCGGT GACATCGAAG  
 551 ACGAGTTTGA CGAAGACGAA AGCGcgcgacg acatCCACTC cgTTTccgCC  
 601 GAACGCTGGC GCATCCacgc ggctaCCGAA ATCGAAGaca TCAACGCCTT  
 651 TTTCCGTACG GAatacggca gcgaagaagc cgacaccatc cggcggctTG  
 701 GTCATTGAGC AATTGGGACA CCTGCCCGTG CGCGGCGAAA AAGTCCTTAT  
 751 cggcggGTTTG Cagttcaccg tCGCCCGCGC CGACAACCGC CGCCTGCACA  
 801 CGCTGATGGC GACCCGCGTG AAGTAAGCAG AGCCTGCCcg AccgcgcttT



851 CTGCacAGTT TAGGatgACG gtaCGGTCGT TTTCTGTTC AATCCGCCCC  
901 ATCCGCCAAA CATAA

This encodes a protein having amino acid sequence <SEQ ID 28; ORF5ng-1>:

```

5      1 MDGAQPKTNF FERLIARLAR EPDSAEDVLN LLRQAHEQEV FDADTLTRLE
      51 KVLDFAELEV RDAMITSRM NVLKENDSIE RITAYVIDTA HSRFPVIGED
101    KDEVLGILHA KDLLKYMFP EQFHLKSVLR PAVFVPEGKS LTALLKEFRE
151    QRNHMAIVID EYGGTSGLVT FEDIIEQIVG DIEDEFDEDE SADDIHSVSA
201    ERWRIHAATE IEDINAFGT EYGSEADTI RRLGHSGIGT PARARRKSPY
251    RRFVHRRPR RQPPAHADG DPREVSRACP TAVSAQFRMT VRSFSVSIRP
10     301 IRQT*

```

The originally-identified partial strain B sequence (ORF5) shows 83.1% identity over a 135aa overlap with the partial gonococcal sequence (ORF5ng):

```

      orf5                                NHMAIVIDEYGGTSGLVTFEDIIEQIVGEI 30
15     orf5ng      FHLKSVLRPAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGLVTFEDIIEQIVGDI 182
      orf5      EDEFDEDDSDADNIHAVSSDTWRIHAATEIEDINTFFGTEYSIEEADTIXRPGHSRVGTSA 90
20     orf5ng      EDEFDEDESADDIHSVSAERWRIHAATEIEDINAFGT EYGSEADTIRRLGHSGIGTPA 242
      orf5      RARRKSPYRRFAVHRRTRRQPPPAYADGDPREVSX----RRFCTV 131
      orf5ng      RARRKSPYRRFAVHRRPRRQPPAHADGDPREVSACPHRRFCTV 287

```

The complete strain B and gonococcal sequences (ORF5-1 & ORF5ng-1) show 92.4% identity in 304 aa overlap:

```

      10      20      30      40      50      60
30     orf5ng-1.pep MDGAQPKTNFFERLIARLAREPDSAEDVLNLLRQAHEQEVFDADTLTRLEKVLDFAELEV
      orf5-1      MDGAQPKTNFFERLIARLAREPDSAEDVLNLLRQAHEQEVFDADTLTRLEKVLDFSDLEV
      10      20      30      40      50      60
35     orf5ng-1.pep RDAMITSRMNVLKENDSIERITAYVIDTAHSRFPVIGEDKDEVLGILHAKDLLKYMFP
      orf5-1      RDAMITSRMNVLKENDSIERITAYVIDTAHSRFPVIGEDKDEVLGILHAKDLLKYMFP
      70      80      90      100     110     120
40     orf5ng-1.pep EQFHLKSVLRPAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGLVTFEDIIEQIVG
      orf5-1      EQFHLKSVLRPAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGLVTFEDIIEQIVG
      130     140     150     160     170     180
45     orf5ng-1.pep DIEDEFDEDESADDIHSVSAERWRIHAATEIEDINAFGT EYGSEADTIRRLGHSGIGT
      orf5-1      EIEDEFDEDDSDADNIHAVSSERWRIHAATEIEDINTFFGTEYSSEEADTIRP-GHSRVGT
      190     200     210     220     230
50     orf5ng-1.pep PARARRKSPYRRFAVHRRPRRQPPAHADGDPREVSACPTAVSAQFRMTVRSFSVSIRP
      orf5-1      SARARRKSPYRRFAVHRRTRRQPPPAYADGDPREVS----TAVSAQFRMTVRAFSVSIRP
      240     250     260     270     280     290
55     orf5ng-1.pep IRQTX
      orf5-1      IRQTX
      300
60

```

Computer analysis of these amino acid sequences indicates a putative leader sequence, and identified the following homologies:

Homology with hemolysin homolog TlyC (accession U32716) of *H. influenzae*

ORF5 and TlyC proteins show 58% aa identity in 77 aa overlap (BLASTp).

```

5      ORF5   2  HMAIVIDEYGGTSGGLVTFEDIIEQIVGEIEDEFDEDDSDADNIHAVSSDTWRIHAATEIED 61
      TlyC   166 HMAIV+DE+G SGLVT EDI+EQIVG+IEDEFDE++ AD I +S T+ + A T+I+D
      HMAIVVDEFGAVSGLVTIEDILEQIVGDIEDEFDEEEIAD-IRQLSRHTYAVRALTDIDD 224

10     ORF5   62 INTFFGTEYSIEEADTI 78
      TlyC   225 FNAQFNTDFDDEEVDTI 241
      N F T++ EE DTI

```

ORF5ng-1 also shows significant homology with TlyC:

```

15     SCORES      Init1:   301 Initn:   419 Opt:    668
      Smith-Waterman score: 668;    45.9% identity in 242 aa overlap

      10          20          30          40          50
or5ng-1.pep  MDGAQPKTNFFERLIARLAR-EPDSAEDVLNLLRQAHEQEVFDADTLTRLEK
      | | : | : | : | : | : | : | : | : | : | : | : | : | : |
20     tlyc_haein  MNDEQQNSNQSENTKKPFFQSLFRFFQGLKNREELVEVIRDSEQNDLIDQNTREMI
      10          20          30          40          50          60

      60          70          80          90          100          109
or5ng-1.pep  VLDFAELEVRDAMITRSRMNVLKENDSIERITAYVIDTAHSRFPVIGE--DKDEVLGILH
      | : : | : | : | : | | | : : : : : : : : : | : | : | : | : | : |
25     tlyc_haein  VMEIAELRVRDIMIPRSQIIFIEDQQDLNTCLNTIESAHSRFPVIADADDRDNVIGILH
      70          80          90          100          110          120

      110          120          130          140          150          160
or5ng-1.pep  AKDLLKYM-FNPEQFHLKSVLRPAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGL
      | : : | : | : | : | | : | : | : | : | : | : | : | : | : | : |
30     tlyc_haein  AKDLLKFLREDAEVDLSSLLRPVVIVPESKRVDRMLKDFRSERFHMMAIVVDEFGAVSGL
      130          140          150          160          170          180

      170          180          190          200          210          220
or5ng-1.pep  VTFEDIIEQIVGDIEDEFDEDESADDIHVSVAERWRIHAATEIEDINAFVGTGTEYGSSEAD
      | : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
35     tlyc_haein  VTIEDILEQIVGDIEDEFDEEEIAD-IRQLSRHTYAVRALTDIDDFNAQFNTDFDDEEVD
      190          200          210          220          230

      230          240          250          260          270          280
or5ng-1.pep  TIRRLGHSGIG-TPARARRKSPYRRFAVHRRPRRQPPAHADGDPREVSRACTAVSAQF
      | | | : : | : | : | : | : | : | : | : | : | : | : | : | : |
40     tlyc_haein  TIGGLIMQTFGYLPKRGEIILKNLQFKVTSADSRRLIQLRVTPDHLAEMNNVDEKSE
      240          250          260          270          280          290

```

Homology with a hypothetical secreted protein from *E. coli*:

ORF5a shows homology to a hypothetical secreted protein from *E. coli*:

```

50     sp|P77392|YBEX_ECOLI HYPOTHETICAL 33.3 KD PROTEIN IN CUTE-ASNB INTERGENIC REGION
      >gi|1778577 (U82598) similar to H. influenzae [Escherichia coli] >gi|1786879
      (AE000170) f292; This 292 aa ORF is 23% identical (9 gaps) to 272 residues of an
      approx. 440 aa protein YTFL_HAEIN SW: P44717 [Escherichia coli] Length = 292

```

```

55     Score = 212 bits (533), Expect = 3e-54
      Identities = 112/230 (48%), Positives = 149/230 (64%), Gaps = 3/230 (1%)

```

```

      Query: 2  DGAQPKTNFXRRLIARLAR-EPDSAEDVLTLLRQAHEQEVFDADTLRLKVLDFSDLEV 60
      D K F L+++L EP + +++L L+R + + ++ D DT LE V+D +D V
      Sbjct: 10 DTISNKKGGFFSLLLSQLFHGEPKNRDELLALIRDSGQNDLIDEDTRDMLEGVMDIADQRV 69

60     Query: 61 RDAMITRSRMNVLKENDSIERITAYVIDTAHSRFPVIGEDKDEVLGILHAKDLLKYM-FN 119

```

RD MI RS+M LK N +++ +I++AHSRFPVI EDKD + GIL AKDLL +M +  
 Sbjct: 70 RDIMIPRSQMITLKRNQTLDDECLDVIIESAHSRFPVISEDKDHIEGILMAKDLLPFMRSD 129

Query: 120 PEQFHLKSILRPAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGSLVTEDIEIQIV 179  
 E F + +LR AV VPE K + +LKEFR QR HMAIVIDE+GG SGLVT EDI+E IV

Sbjct: 130 AEAFSMDKVLQRQAVVVPESKRVDRLKEFRSQRYHMAIVIDEFGGVSGSLVTIEDILELIV 189

Query: 180 GDIEDEFDEDESADNIHAVSAERWRIHAATEIEDINAFFGTEYSSEEADT 229  
 G+IEDE+DE++ D +S W + A IED N FGT +S EE DT

Sbjct: 190 GEIEDEYDEEDDID-FRQLSRHTWTVRALASIEDFNFAFGTHFSDEEVD 238

Based on this analysis, including the amino acid homology to the TlyC hemolysin-homologue from *H. influenzae* (hemolysins are secreted proteins), it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae* are secreted and could thus be useful antigens for vaccines or diagnostics.

ORF5-1 (30.7kDa) was cloned in the pGex vector and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 2A shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein was used to immunise mice, whose sera were used for Western blot analysis (Figure 1B). These experiments confirm that ORF5-1 is a surface-exposed protein, and that it is a useful immunogen.

## Example 5

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 29>:

1 ATGCGCGGCG GCAGGCCGGA TTCCGTTACC GTGCAGATTA TCGAAGGTTG  
 51 GCGTTTTTCG CATATGAGGA AAGTCATCGA CGCAACGCCG GACATCGGAC  
 101 ACGACACCAA AGGCTGGAGC AATGAAAAAC TGATGGCGGA AGTTGCGCCC  
 151 GATGCCTTCA GCGGCAATCC TGAAGGGCAG TTTTCCCCCG ACAGCTACGA  
 201 AATCGATGCG GCGGCGAGTG ATTGCAGAT TTACCAAACC GCCTACAAGG  
 251 GCGATGCAAC GCCGCCTGAA TGAAGGCATG GGAAAGCAGG CAGGACGGGC  
 301 TGCCTTATAA AAACCTTAT GAAATGCTGA TTATGGCGAG CCTGGTCGAA  
 351 AAGGAAACAG GGCATGAAGC CGAsCsGAC CATGTcGCTT CCGTCTTCGT  
 401 CAACCGCCTG AAAATCGGTA TCGCCTGCA AACCGAssCG TCCGTGATTT  
 451 ACGGCATGGG TCGGCATAC AAGGGCAAAA TCCGTAAAGC CGACCTGCGC  
 501 CGCGACACGC CGTACAACAC CTACACGCGC GCGGTCTGCG CGCCAACCCC  
 551 GATTGCGCTG CCC..

This corresponds to the amino acid sequence <SEQ ID 30; ORF7>:

1 MRGGRPDSVT VQIEGSRFS HMRKVIDATP DIGHDTKGWS NEKLMAEVAP  
 51 DAFSGNPEGQ FFPDSYEIDA GGSDLQIYQT AYKAMQRRLN EAWESRQDGL  
 101 PYKNPYEMLI MAXLVEKETG HEAXXDHVAS VFNRLKIGM RLQTXSVIY  
 151 GMGAAYKGKI RKADLRDTP YNTYTRGGLP PTPIALP..

Further sequence analysis revealed the complete DNA sequence <SEQ ID 31>:

1 ATGTTGAGAA AATTGTTGAA ATGGTCTGCC GTTTTTTTGA CCGTGTCCGGC  
 51 AGCCGTTTTC GCCGCGCTGC TTTTGTGTC TAAGGATAAC GGCAGGGCAT  
 101 ACCGAATCAA AATTGCCAAA AACCAGGGTA TTTCGTCCGT CGGCAGGAAA  
 151 CTTGCCGAAG ACCGCATCGT GTTCAGCAGG CATGTTTTGA CGGCGGCGGC  
 201 CTACGTTTTG GGTGTGCACA ACAGGCTGCA TACGGGGACG TACAGATTGC  
 251 CTTCCGAAGT GTCTGCTTGG GATATCTTGC AGAAAATGCG CGGCGGCAGG  
 301 CCGGATTCCG TTACCGTGCA GATTATCGAA GGTTCGCGTT TTTTCGATAT  
 351 GAGGAAAGTC ATCGACGCAA CGCCCGACAT CGGACACGAC ACCAAAGGCT  
 401 GGAGCAATGA AAAACTGATG GCGGAAGTTG CGCCCGATGC CTTACGCGGC  
 451 AATCCTGAAG GGCAGTTTTT CCCCACAGC TACGAAATCG ATCGGGCGGC

5

10

501	CAGTGATTTG	CAGATTTACC	AAACCGCCTA	CAAGGCGATG	CAACGCCGCC
551	TGAATGAGGC	ATGGGAAAGC	AGGCAGGACG	GGTGCCTTA	TAAAAACCCT
601	TATGAAATGC	TGATTATGCG	GAGCCTGGTC	GAAGAAGAAA	CAGGGCATGA
651	AGCCGACCGC	GACCATGTGC	CTTCCGTCTT	CGTCAACCGC	CTGAAAAATCG
701	GTATGCGCCT	GCAAACCGCA	CCGTCCGTGA	TTTACGGCAT	GGGTCCGGCA
751	TACAAGGGCA	AAATCCGTAA	AGCCGACCTG	CGCCGCGACA	CGCCGTACAA
801	CACCTACACG	CGCGGCGGTC	TGCCGCCAAC	CCCGATTGCG	CTGCCCGGCA
851	AGGCGGCACT	CGATGCCGCC	GCCCATCCGT	CCGGCGAAAA	ATACCTGTAT
901	TTCGTGTCCA	AAATGGACGG	CACGGGCTTG	AGCCAGCTTCA	GCCATGATT
951	GACCGAACAC	AAATGCCGCG	TCCGCAATA	TATTTTGAAA	AAATAA

This corresponds to the amino acid sequence <SEQ ID 32; ORF7-1>:

1	MLRKLLKWSA	VFLTVSAAVF	AALLFVPKDN	GRAYRIKIAK	NQGISSVGRK
51	LAEDRIVFSR	HVLTAAYVL	GVHNLHTGT	YRLPSEVSAW	DILQKMRGRG
101	PDSVTVQIIE	GSRFSHMRKV	IDATPDIGHD	TKGWSNEKLM	AEVAPDAFSG
151	NPEGQFFPDS	YEIFAGGSDL	QIYQTAYKAM	QRRLENAEWES	RQDGLPYKPN
201	YEMLIASLV	EKETGHEADR	DHVASVFVNR	LKIGMRLQTD	PSVIYGMGAA
251	YKGKIRKADL	RRDTPYNTY	RGGLPPTPIA	LPKGAALDAA	AHPSGEKYLY
301	FVSKMDGTGL	SQFSDHLTEH	NAAVRKYILK	K*	

Computer analysis of this amino acid sequence gave the following results:

20 Homology with hypothetical protein encoded by *yceg* gene (accession P44270) of *H.influenzae*  
ORF7 and yceg proteins show 44% aa identity in 192 aa overlap:

25	ORF7	1	MRGGRPDSVTVQIIIEGSRFSHMRKVIDATPDIGHDTKGWSNEKLMA-----EVAPDAFSG	55
	yceg	102	LNSGKEVQFNVKWIEGKTFKDWKDLLENAPHLVQTLKDKSNEEIFALLDLPDIGQNLELK	161
30	ORF7	56	NPEGQFFPDSYEIDAGGSDLQIYQTAYKAMQRRLNEAWESRODGLPYKNPYEMLIMAXLV	115
	yceg	162	NVEGWLYPDTYNYTPKSTDLELLKRS AERMKKALNKAWNERDEDLPLANPYEMLILASIV	221
35	ORF7	116	EKETGHEAXXDHVASVFVNRLKIGMRLQTXXSVIYMGGAAYKGKIRKADLRDTPYNTYT	175
	yceg	222	EKETGIANERAKVASVFINRLKAKMKLQTDPTVIYMGGENYNGNIRKKDLETKTPYNTYV	281
35	ORF7	176	RGLPPTPIALP	187
	yceg	282	IDGLPPTPIAMP	293

The complete length YCEG protein has sequence:

```

1  MKKFLLIAILL LILILAGVAS FSYKMTFV KTPVNVQADE LLTIERGTTT
40 51  SKLATLFEQE KLIADGKLLP YLLKLPKPELN KIKAGTYSLE NVKTVQDLLD
101 LLNSGKEVQF NVKVGIEGTF KDWKRDLENA PHLVQTLKDK SNEEIFALLD
151 LPDIGNQLLEL KNEVGWLYPD TYNYPKSTD LELKRSARA MKKALNKAWN
201 ERDEDLPLAN PYEMLILASI VEKETGIANE RAKVASVFIN RLLKAMKMLQT
251 DPTVIYGMGE NYNGNIRKKD LETKTPYNTY VIDGLPPTPI AMPSESSLQA
301 VANPEKTDYF YFVADGSGGH KETRNLNEHN KAVEYLRWY RSQKNK

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF7 shows 95.2% identity over a 187aa overlap with an ORF (ORF7a) from strain A of *N. meningitidis*:

```

50      orf7.pep      10      20      30
                        MRGGRPDSVTVQIIIEGSRFSHMRKVIDATP
                        |||
orf7a      AAYVLGVHNRRLHTGTYRLPSEVSAWDILQMRGGRPDSVTVQIIIEGSRFSHMRKVIDATP
              70      80      90      100      110      120

55      orf7.pep      40      50      60      70      80      90
DIGHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLQIYQTAYKAMQRRLN
|||

```

orf7a	DIEHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLRIYQIAYKAMQRRLN
	130 140 150 160 170 180
5	orf7.pep
	100 110 120 130 140 150
	EAWESRQDGLPYKNPYEMLIMAXLVEKETGHEAXXDHVASVFVNRLKIGMRLQTXSVIY
	orf7a
	EAWESRQDGLPYKNPYEMLIMASLIEKETGHEADRDHVASVFVNRLKIGMRLQTDPSVIY
	190 200 210 220 230 240
10	orf7.pep
	160 170 180
	GMGAAYKGKIRKADLRDTPYNTYTRGGLPPTPIALP
	orf7a
	GMGAAYKGKIRKADLRDTPYNTYTRGGLPPTPIALPGKAALDAAAHPSGEKYLYFVSKM
	250 260 270 280 290 300
15	orf7a
	DGTGLSQFSHDLTEHNAAVRKYLKX
	310 320 330

The complete length ORF7a nucleotide sequence <SEQ ID 33> is:

20	1	ATGTTGAGAA	AATTGTTGAA	ATGGTCTGCC	GTTTTTTTGA	CCGTATCGGC
	51	AGCCGTTTTT	GCCGCGTCTC	TTTTCGTCCC	TAAAGACAAC	GGCAGGGCAT
	101	ACAGGATTAA	AATTGCCAAA	AACCAGGTA	TTTCGTCCGT	CGGCAGGAAA
	151	CTTGCCGAAG	ACCGCATCGT	GTTTCAGCAGG	CATGTTTGA	CGGCGGCGGC
	201	CTACGTTT	GGTGTGCACA	ACAGGCTGCA	TACGGGGACG	TACAGACTGC
	251	CTTCGGAAGT	GTCTGCTTGG	GATATCTTGC	AGAAAATGCG	CGGCGGCGAGG
25	301	CCGGATTCCG	TTACCGTGCA	GATTATCGAA	GGTTCGCGTT	TTTCGCATAT
	351	GAGGAAAGTC	ATCGACGCAA	CGCCCGACAT	CGAACACGAC	ACCAAAGGCT
	401	GGAGCAATGA	AAAACATGAT	GCGGAAGTTG	CCCCTGATGC	CTTCAGCGGC
	451	AATCCTGAAG	GGCAGTTTTT	CCCCGACAGC	TACGAAATCG	ATGCGGGCGG
	501	CAGCGATTTA	CGGATTTACC	AAATCGCCTA	CAAGGCGATG	CAACGCCGAC
30	551	TGAATGAGGC	ATGGGAAAGC	AGGCAGGACG	GGCTGCCTTA	TAAAAACCTT
	601	TATGAAATGC	TGATTATGGC	GAGCCTGATC	GAAAAGGAAA	CAGGGCATGA
	651	AGCCGACCGC	GACCATGTCG	CTTCCGTCTT	CGTCAACCGC	CTGAAAATCG
	701	GTATGCGCCT	GCAAACCGAC	CCGTCCGTGA	TTTACCGCAT	GGGTGCCGCA
35	751	TACAAGGGCA	AAATCCGTAA	AGCCGACCTG	CGCCGCGACA	CGCCGTACAA
	801	CACCTACACG	CGCGGCGGTC	TGCCGCCAAC	CCCGATCGCG	CTGCCCGGCA
	851	AGGCGGCACT	CGATGCCGCC	GCCCATCCGT	CCGGTAAAAA	ATACCTGTAT
	901	TTCTGTCCA	AAATGGACGG	TACGGGCTTG	AGCCAGTTCA	GCCATGATTT
	951	GACCGAACAC	AACGCCGCCG	TTTCGCAATA	TATTTTGAAA	AAATAA

This is predicted to encode a protein having amino acid sequence <SEQ ID 34>:

40	1	<u>MLRKLLKWSA</u>	<u>VFLTVSAAVF</u>	<u>AALLFVPKDN</u>	GRAYRIKIAK	NQGISSVGRK
	51	<u>LAEDRIVFSR</u>	<u>HVLTAAAYVL</u>	<u>GVHNRHLTGT</u>	YRLPSEVSAW	DILQKMRGGR
	101	<u>PDSVTVQIIE</u>	<u>GSRFSHMRKV</u>	<u>IDATPDIEHD</u>	TKGWSNEKLM	AEVAPDAFSG
	151	<u>NPEGQFFPDS</u>	<u>YEIDAGGSDL</u>	<u>RIYQIAYKAM</u>	QRRLEAWES	RQDGLPYKNP
45	201	<u>YEMLIMASLI</u>	<u>EKETGHEADR</u>	<u>DHVASVFVNR</u>	LKIGMRLQTD	PSVIYGMGAA
	251	<u>YKGKIRKADL</u>	<u>RRDTPYNTYT</u>	<u>RGGLPPTPIA</u>	LPGKAALDAA	AHPSGEKYL
	301	<u>FVSKMDGTGL</u>	<u>SQFSHDLTEH</u>	<u>NAAVRKYLK</u>	K*	

A leader peptide is underlined.

ORF7a and ORF7-1 show 98.8% identity in 331 aa overlap:

50	orf7a.pep	10 20 30 40 50 60
	MLRKLLKWSAVFLTVSAAVF AALLFVPKDN GRAYRIKIAKNQGISSVGRKLAEDRIVFSR	
	orf7-1	10 20 30 40 50 60
	MLRKLLKWSAVFLTVSAAVF AALLFVPKDN GRAYRIKIAKNQGISSVGRKLAEDRIVFSR	
55	orf7a.pep	70 80 90 100 110 120
	HVLTAAAYVLGVHNRHLTGT YRLPSEVSAWDILQKMRGGRPDSVTVQIIEGSRFSHMRKV	
	orf7-1	70 80 90 100 110 120
	HVLTAAAYVLGVHNRHLTGT YRLPSEVSAWDILQKMRGGRPDSVTVQIIEGSRFSHMRKV	
60		130 140 150 160 170 180

	orf7a.pep	IDATPDIEHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLRIYQIAYKAM
	orf7-1	IDATPDIGHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLQIYQTAYKAM
5		130 140 150 160 170 180
	orf7a.pep	190 200 210 220 230 240
	orf7-1	QRRLEAWESRQDGLPYKNPYEMLIMASLIEKETGHEADRDHVASVFNRLKIGMRLQTD
10		190 200 210 220 230 240
	orf7a.pep	250 260 270 280 290 300
	orf7-1	PSVIYGMGAAYKGKIRKADLRDTPYNTYTRGGLPPTPIALPGKAALDAAAHPSGEKYL
15		250 260 270 280 290 300
	orf7a.pep	310 320 330
	orf7-1	FVSKMDGTGLSQFSHDLTEHNAAVRKYLKXX
20		310 320 330

### Homology with a predicted ORF from *N.gonorrhoeae*

25 ORF7 shows 94.7% identity over a 187aa overlap with a predicted ORF (ORF7.ng) from *N. gonorrhoeae*:

	orf7	MRGGRPDVSVTVQIIIEGSRFSHMRKVIDATPDIGHDTKGWSNEKLMAEVAPDAFSGNPEGQ	60
	orf7ng	MRGGRPDVSVTVQIIIEGSRFSHMRKVIDATPDIGHDTKGWSNEKLMAEVAPDAFSGNPEGQ	60
30	orf7	FFPDSYEIDAGGSDLQIYQTAYKAMQRRLEAWESRQDGLPYKNPYEMLIMASLVEKETG	120
	orf7ng	FFPDSYEIDAGGSDLQIYQTAYKAMQRRLEAWAGRQDGLPYKNPYEMLIMASLIEKETG	120
35	orf7	HEAXXDHVASVFNRLKIGMRLQTXSVIYGMGAAYKGKIRKADLRDTPYNTYTRGGLP	180
	orf7ng	HEADRDHVASVFNRLKIGMRLQTDPSVIYGMGAAYKGKIRKADLRDTPYNTYTGGGLP	180
40	orf7	PTPIALP	187
	orf7ng	PTPIALPGKAAMDAAAHPSGEKYL FVSKMDGTGLSQFSHDLTEHNAAVRKYLKX	236

An ORF7ng nucleotide sequence <SEQ ID 35> is predicted to encode a protein having amino acid sequence <SEQ ID 36>:

45	1	MRGGRPDVSVT	VQIIIEGSRFS	HMRKVIDATP	DIGHDTKGWS	NEKLMAEVAP
	51	DAFSGNPEGQ	FFPDSYEIDA	GGSDLQIYQT	AYKAMQRRLN	EAWAGRQDGL
	101	PYKNPYEMLI	MASLIEKETG	HEADRDHVAS	VFNRLKIGM	RLQTDPSVIY
	151	GMGAAYKGKI	RKADLRDTP	YNTYTGGGLP	PTPIALPGKA	AMDAAHPSG
	201	EKYL FVSKM	DGTGLSQFSH	DLTEHNAAVR	KYLKX*	

Further sequence analysis revealed a partial DNA sequence of ORF7ng <SEQ ID 37>:

50	1	..taccgaatca	AGATTGCCAA	AAATCAGGGT	ATTCGTCGG	TCGGCAGGAA
	51	ACTTGCcgaA	GACCGCATCG	TGTT CAGCAG	GCATGTTTG	ACAGCGGCGG
	101	CCTACGTTTT	GGGTGTGCAC	AACAGGCTGC	ATACGGGGAC	gTACAGATTG
	151	CCTTCGGAAG	TGCTGCTTG	GGATATCTTG	CAGAAAATGC	GCGGCGGCAG
55	201	GCCGGATTCC	GTTACCGTGC	AGATTATCGA	AGGTTTCGCGT	TTTTCGCATA
	251	TGAGGAAAGT	CATCGACGCA	ACGCCCGACA	TCGGACACGA	CACCAAAGGC
	301	TGGAGCAATG	AAAAACTGAT	GGCGGAAGTT	GCGCCCGATG	CCTTCAGCGG
	351	CAATCCTGAA	GGGCAGTTTT	TTCGCGACAG	CTACGAAATC	GATGCGGGCG
	401	GCAGCGATTT	GCAGATTAC	CAAACCGCCT	ACAAGGCGAT	GCAACGCCGC
	451	CTGAACGAGG	CATGGGCAGG	CAGGCAGGAC	GGGCTGCCTT	ATAAAAACCC
60	501	TTATGAAATG	CTGATTATGG	CGAGCCTGAT	CGAAAAGGAA	ACGGGGCATG

551 AGGCCGACCG CGACCATGTC GCTTCGGTCT TCGTCAACCG CCTGAAAATC  
 601 GGTATGCGCC TGCAAACCGA CCCGTCCGTG ATTTACGGCA TGGGTGCGGC  
 651 ATACAAGGGC AAAATCCGTA AAGCCGACCT GCGCCGCGAC ACGCCGTACA  
 701 aCAccTAtac gggcgggggc ttgccgccaa cccggattgc gctgcccggc  
 751 Aaggcgggcaa tggatgccgc cgcccaccgc tccggcgaAa aatacctgTa  
 801 tttcgtgtcC AAAATGGACG GCACGGGCTT GAGCCAGTTC AGCCATGATT  
 851 TGACCGAACA CAACGCCGCC gTcCGCAAAT ATATTTTGAA AAAATAA

This corresponds to the amino acid sequence <SEQ ID 38; ORF7ng-1>:

1 .YRIKIAKNQG ISSVGRKLAE DRIVFSRHVL TAAAYVLGVH NRLHTGTYRL  
 51 PSEVSAWDIL QKMRGGRPDS VTVQIIIEGSR FSHMRKVIDA TPDIGHDTKG  
 101 WSNEKLMAEV APDAFSGNPE GQFFPDSYEI DAGGSDLQIY QTAYKAMQRR  
 151 LNEAWAGRQD GLPYKNPYEM LIMASLIEKE TGHEADRDHV ASVFNRLKI  
 201 GMRLQTDPSV IYGMGAAYKG KIRKADLRD TPYNTYTGGG LPPTRIALPG  
 251 KAAMDAAHP SGEKYLIFVS KMDGTGLSQF SHDLTEHNAA VRKYILKK\*

15 ORF7ng-1 and ORF7-1 show 98.0% identity in 298 aa overlap:

10 orf7-1.pep 10 20 30 40 50 60  
 KLLKWSAVFLTVSAAVFAALLFVPKDNGRAYRIKIAKNQGISSVGRKLAEDRIVFSRHVL  
 orf7ng-1 YRIKIAKNQGISSVGRKLAEDRIVFSRHVL  
 20 10 20 30  
 25 orf7-1.pep 70 80 90 100 110 120  
 TAAAYVLGVHNRLHTGTYRLPSEVSAWDILQKMRGGRPDSVTVQIIIEGSRFSHMRKVIDA  
 orf7ng-1 TAAAYVLGVHNRLHTGTYRLPSEVSAWDILQKMRGGRPDSVTVQIIIEGSRFSHMRKVIDA  
 40 50 60 70 80 90  
 30 orf7-1.pep 130 140 150 160 170 180  
 TPDIGHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLQIYQTAYKAMQRR  
 orf7ng-1 TPDIGHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLQIYQTAYKAMQRR  
 100 110 120 130 140 150  
 35 orf7-1.pep 190 200 210 220 230 240  
 LNEAWESRQDGLPYKNPYEMLIMASLVEKETGHEADRDHVASVFNRLKIGMRLQTDPSV  
 orf7ng-1 LNEAWAGRQDGLPYKNPYEMLIMASLIEKETGHEADRDHVASVFNRLKIGMRLQTDPSV  
 160 170 180 190 200 210  
 40 orf7-1.pep 250 260 270 280 290 300  
 IYGMGAAYKGKIRKADLRDTPYNTYTRGGLPPTPIALPGKAALDAAAHPSGEKYLIFVS  
 orf7ng-1 IYGMGAAYKGKIRKADLRDTPYNTYTGGGLPPTRIALPGKAAMDAAAHPSGEKYLIFVS  
 220 230 240 250 260 270  
 45 orf7-1.pep 310 320 330  
 KMDGTGLSQFSHDLTEHNAAVRKYILKKX  
 orf7ng-1 KMDGTGLSQFSHDLTEHNAAVRKYILKKX  
 280 290

In addition, ORF7ng-1 shows significant homology with a hypothetical *E.coli* protein:

55 sp|P28306|YCEG\_ECOLI HYPOTHETICAL 38.2 KD PROTEIN IN PABC-HOLB INTERGENIC REGION  
 gi|1787339 (AE000210) o340; 100% identical to fragment YCEG\_ECOLI SW: P28306 but  
 has 97 additional C-terminal residues [Escherichia coli] Length = 340  
 Score = 79 (36.2 bits), Expect = 5.0e-57, Sum P(2) = 5.0e-57  
 Identities = 20/87 (22%), Positives = 40/87 (45%)  
 60 Query: 10 GISSVGRKLAEDRIVFSRHVL TAAAYVLGVHNRLHTGTYRLPSEVSAWDILQKMRGGRPD 69  
 G ++G +L D+I+ V + + GTYR +++ ++L+ + G+  
 Sbjct: 49 GRLALGEQLYADKIINRPRVFQWLLRIEPLDLSHFAGTYRFTPTQMTVREMLKLLSESGKEA 108  
 Query: 70 SVTVQIIIEGSRFSHMRKVIDATPDIGH 96

++++EG R S K + P I H  
 Sbjct: 109 QFPLRLVEGMRLSDYLKQLREAPYIKH 135

Score = 438 (200.7 bits), Expect = 5.0e-57, Sum P(2) = 5.0e-57  
 Identities = 84/155 (54%), Positives = 111/155 (71%)

5 Query: 120 EGQFFPDSEYIDAGGSDLQIYQTAYKAMQRRRLNEAWAGRQDGLPYKNPYEMLIMASLIEK 179  
 EG F+PD++ A +D+ + + A+K M + ++ AW GR DGLPYK+ +++ MAS+IEK  
 10 Sbjct: 158 EGWFWDPTWMYTANTTDDVALLKRAHKMKVAVDSAWEGRADGLPYKDKNQLVTMASIIEK 217

Query: 180 ETGHEADRDHVASVFVNRLKIGMRLQTDPSVIYGMGAAYKGKIRKADLRDTPYNTYTGG 239  
 ET ++RD VASVF+NRL+IGMRLQTD+VIYGMG Y GK+ +ADL T YNTYT  
 15 Sbjct: 218 ETAVASERDKVASVFINRLRIGMRLQTDPTVIYGMGERYNGKLSRADLETPTAYNTYTIT 277

Query: 240 GLPPTRIALPGKAAMDAAAHPSGEKYLYFVSKMDG 274  
 GLPP IA PG ++ AAAHP+ YLYFV+ G  
 15 Sbjct: 278 GLPPGAIATPGADSLKAAAHPAKTPYLYFVADGKG 312

Based on this analysis, including the fact that the *H.influenzae* YCEG protein possesses a possible  
 20 leader sequence, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their  
 epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 6

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 39>:

25 1 CGTTTCAAAA TGTTAACTGT GTTGACGGCA ACCTTGATTG CCGGACAGGT  
 51 ATCTGCCCGC GGAGGCGGTG CCGGGGATAT GAAACAGCCG AAGGAAGTCG  
 101 GAAAGGTTT CAGAAAGCAG CAGCGTTACA GCGAGGAAGA AATCAAAAAC  
 151 GAACGCGCAC GGCCTGCGGC AGTGGGCGAG CGGGTTAATC AGATATTAC  
 201 GTTGCTGGGA GGGGAAACCG CCTTGCAAAA GGGGACGGCG GGAACGGCTC  
 251 TGGCAACCTA TATGCTGATG TTGGAACGCA CAAAATCCCC CGAAGTCGCC  
 30 301 GAACGCGCCT TGGAAATGGC CGTGTCGCTG AACGCGTTG AACAGGCGGA  
 351 AATGATTAT CAGAAATGGC GGCAGATTGA GCCTATACCG GGTAAAGCGC  
 401 AAAAACGGGC GGGGTGGCTG CGGAACGTGC TGAGGGAAAG AGGAAATCAG  
 451 CATCTGGACG GACGGGAAGA AGTGCTGGCT CAGGCGGACG AAGGACAG

This corresponds to the amino acid sequence <SEQ ID 40; ORF9>:

35 1 .RFKMLTVLTA TLIAGQVSAA GGGAGDMKQP KEVGKVERKQ QRYSEEEIKN  
 51 ERARLAAVGE RVNQIFTLG GETALQKQA GTALATYMLM LERTKSPEVA  
 101 ERALEMAVSL NAFEQAEMII QKWRQIEPIP GKAQKRAGWL RNVLRERNGQ  
 151 HLDGREEVLA QADEGQ

Further sequence analysis revealed the complete DNA sequence <SEQ ID 41>:

40 1 ATGTTACCTA ACCGTTTCAA AATGTTAACT GTGTTGACGG CAACCTTGAT  
 51 TGCCGGACAG GTATCTGCCG CCGGAGGCGG TGCGGGGAT ATGAAACAGC  
 101 CGAAGGAAGT CGGAAAGGTT TTCAGAAAGC AGCAGCGTTA CAGCGAGGAA  
 151 GAAATCAAAA ACGAACGCGC ACGGCTTGCG GCAGTGGGCG AGCGGGTTAA  
 201 TCAGATATTT ACGTTGCTGG GAGGGGAAAC CGCCTTGCAA AAGGGGACAG  
 45 251 CGGGAACGCG TCTGGCAACC TATATGCTGA GTTGGAACG CACAAAATCC  
 301 CCCGAAGTCG CCGAACGCGC CTTGGAATG GCCGTGTCGC TGAACGCGTT  
 351 TGAACAGGCG GAAATGATT ATCAGAAATG GCGGCAGATT GAGCCTATAC  
 401 CCGGTAAGGC GCAAAAACGG CCGGGGTGGC TGCGGAACGT GCTGAGGGAA  
 451 AGAGGAAATC AGCATCTGGA CGGACTGGAA GAAGTGCTGG CTCAGGCGGA  
 50 501 CGAAGGACAG AACCGCAGGG TGTTTTATT GTTGGCACA GCGCCGTGC  
 551 AACAGGACGG GTTGGCGCAA AAAGCATCGA AAGCGGTTCT CCGCGCGGCG  
 601 TTGAAATATG AACATCTGCC CGAAGCGGCG GTTGCCGATG TGGTGTTTCA  
 651 CGTACAGGGA CGCGAAAAGG AAAAGGCAAT CGGAGCTTTG CAGCGTTTGG  
 701 CGAAGCTCGA TACGAAATA TTGCCCCCA CTTTAATGAC GTTGCCTCTG  
 751 ACTGCACGCA AATATCCCGA AATACTCGAC GGCTTTTTCG AGCAGACAGA  
 801 CACCAAAAAC CTTTCGGGCG TCTGGCAGGA AATGGAAATT ATGAATCTCG  
 851 TTTCCCTGCA CAGGCTGGAT GATGCCTATG CGCGTTTGAA CGTGCTGTTG



5  
10  
15  
20

```

901 GAACGCAATC CGAATGCAGA CCTGTATATT CAGGCAGCGA TATTGGCGGC
951 AAACCGAAAA GAAGGTGCTT CCGTTATCGA CGGCTACGCC GAAAAGGCAT
1001 ACGGCAGGGG GACGGAGGAA CAGCGGAGCA GGGCGGCGCT AACGGCGCGC
1051 ATGATGTATG CCGACCGCAG GGATTACGCC AAAGTCAGGC AGTGGCTGAA
1101 AAAAGTATCC GCGCCGGAAT ACCTGTTCGA CAAAGGTGTG CTGGCGGCTG
1151 CGGCGGCTGT CGAGTTGGAC GCGCGCAGGG CGGCTTTGCG GCAGATCGGC
1201 AGGGTGCGGA AACTTCCCGA ACAGCAGGGG CGGTATTTTA CGGCAGACAA
1251 TTTGTCCAAA ATACAGATGC TCGCCCTGTC GAAGCTGCCC GATAAACGGG
1301 AGGCTTTGAG GGGGTGGAC AAGATTATCG AAAAACCGCC TGCCGGCAGT
1351 AATACAGAGT TACAGGCAGA GGCATTGGTA CAGCGGTCAG TTGTTTACGA
1401 TCGGCTTGGC AAGCGGAAAA AAATGATTTC AGATCTGAA AGGGCGTTCA
1451 GGCTTGACCC CGATAACGCT CAGATTATGA ATAATCTGGG CTACAGCCTG
1501 CTGACCGATT CCAAACGTTT GGACGAAGGT TTCGCCCTGC TTCAGACGGC
1551 ATACCAAATC AACCCGGACG ATACCGCTGT CAACGACAGC ATAGGCTGGG
1601 CGTATTACCT GAAAGGCGAC GCGGAAAGCG CGCTGCCGTA TCTGCGGTAT
1651 TCGTTTGAAG ACGACCCCGA GCCCGAAGTT GCCGCCCAT TGGGCGAAGT
1701 GTTGTGGGCA TTGGGCGAAC GCGATCAGGC GGTTGACGTA TGGACGACG
1751 CGGCACACCT TACGGGAGAC AAGAAAATAT GGCGGGAAC GCTCAAACGT
1801 CACGGCATCG CATTGCCCA ACCTTCCGA AAACCTCGGA AATAA

```

20 This corresponds to the amino acid sequence <SEQ ID 42; ORF9-1>:

25  
30  
35

```

1 MLPNRFKMLT VLTATLIAGQ VSAAGGGAGD MKQPKEVGKV FRKQORYSEE
51 EIKNERARLA AVGERVNQIF TLLGGETALQ KGQAGTALAT YMLMLERTKS
101 PEVAERALEM AVSLNAFEQA EMIYQKWRQI EPIPGKAQKR AGWLRNVLRE
151 RGNQHLDGLE EVLAQADEGQ NRRVFLLLAQ AAVQQDGLAQ KASKAVRRAA
201 LKYEHLPEAA VADVFSVQG REKEKAIGAL QRLAKLDTEI LPPTLMTLRL
251 TARKYPEILD GFFEQTDTQN LSAVWQEMEI MNLVSLHRLD DAYARLNVLL
301 ERNPNADLYI QAAILAANRK EGASVIDGYA EKAYGRGTEE QRSRAALTA
351 MMYADRRDYA KVRQWLKKVS APEYLFDKGV LAAAAAVELD GGRAALRQIG
401 RVRKLPEQQG RYFTADNLSK IQMLALSKLP DKREALRGLD KIEKPPAGS
451 NTELQAEALV QRSVVYDRLG KRKKMISDLE RAFRLAPDNA QIMNNLGYSL
501 LTDSKRLDEG FALLQATAYQI NPDDTAVNDS IGWAYYLGKD AESALPYLRY
551 SFENDPEPEV AAHLGEVLWA LGERDQAVDV WTQAAHLTGD KKIWRETLKR
601 HGIALPQPSR KPRK*

```

Computer analysis of this amino acid sequence gave the following results:

35 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF9 shows 89.8% identity over a 166aa overlap with an ORF (ORF9a) from strain A of *N. meningitidis*:

40  
45  
50  
55

```

          10      20      30      40      50
orf9.pep  RFKMLTVLTATLIAGQVSAAGGGAGDMKQPKEVGKVFRKQORYSEEEIKNERARLA
          ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
orf9a     MLPARFTILSVLAAALLAGQAYAA--GAADAKPPKEVGKVFRKQORYSEEEIKNERARLA
          10      20      30      40      50

          60      70      80      90     100     110
orf9.pep  AVGERVNQIFTLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA
          |||||::||::||::||::||::||::||::||::||::||::||::||::||::||
orf9a     AVGERVNQIFTLGXETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA
          60      70      80      90     100     110

          120     130     140     150     160
orf9.pep  EMIYQKWRQIEPIPGKAQKRAGWLRNVLRERGNQHLDGREEVLAQADEGQ
          |||||::||::||::||::||::||::||::||::||::||::||::||::||
orf9a     EMIYQKWRQIEPIPGKAQKRAGWLRNVLRERGNQHLDGLEEXLAQADEXQNRRVFLLLAQ
          120     130     140     150     160     170

          180     190     200     210     220     230
orf9a     AAVQQDGLAQKASKAVRRAALRYEHLPEAAVADVFSVQXREKEKAIGALQRLAKLDTEI

```

The complete length ORF9a nucleotide sequence <SEQ ID 43> is:

60

```

1 ATGTTACCCG CCCGTTTCAC CATTTTATCT GTGCTCGCGG CAGCCCTGCT
51 TGCCGGGCAG GCGTATGCCG CCGGCGCGGC GGATGCGAAG CCGCCGAAGG

```

101 AAGTCGGAAG GGTTCACAGA AAGCAGCAGC GTTACAGCGA GGAAGAAATC  
 151 AAAAACGAAC GCGCACGGCT TGCGGCAGTG GGCGAGCGGG TTAATCAGAT  
 201 ATTTACGTTG CTGGGANGGG AAACCGCCTT GCAAAAGGGG CAGGCGGGAA  
 251 CGGCTCTGGC AACCTATATG CTGATGTTGG AACGCACAAA ATCCCCGAA  
 301 GTCGCCGAAC GCGCCTTGGA AATGGCCGTG TCNCTGAACG CGTTTGAACA  
 351 GGCGGAAATG ATTTATCAGA AATGGCCGGA GATTGAGCCT ATACCGGGTA  
 401 AGGCGCAAAA ACGGGCGGGG TGGCTGCGGA ACGTGCTGAG GGAAGAGGA  
 451 AATCAGCATC TAGACGGAAT GGAAGAANTG CTGGCTCAGG CGGACGAANG  
 501 ACAGAACCGC AGGGTGTGTT TATTGTTGGC ACAAGCCGCC GTGCAACAGG  
 551 ACGGGTTGGC GCAAAAAGCA TCGAAAGCGG TTCGCCGCGC GGCCTTGAGA  
 601 TATGAACATC TGCCCGAAGC GGCGGTTGCC GATGTGGTGT TCAGCGTACA  
 651 GGNACGCGAA AAGGAAAAGG CAATCGGAGC TTTGCAGCGT TTGGCGAAGC  
 701 TCGATACGGA AATATTGCCC CCCACTTTAA TGACGTTGCG TCTGACTGCA  
 751 CGCAAATATC CCGAAATACT CGACGGCTTT TCGAGCAGA CAGACACCCA  
 801 AAACCTTTTCG GCCGTCTGGC AGGAAATGGA AATTATGAAT CTGGTTTCCC  
 851 TGCACAGGCT GGATGATGCC TATGCGCGTT TGACCGTGTG GTTGAACCGC  
 901 AATCCGAATG CAGACCTGTA TATTCAGGCA GCGATATTGG CGGCAACCGC  
 951 AAAAGAANGT GCTTCGTTA TCGACGGCTA CGCCGAAAAG GCATACGGCA  
 1001 GGGGACGCGG GGAACAGCGG GGCAGGCGCG CAATGACGGC GGCATGATA  
 1051 TATGCCGACC GAAGGGATTA CACCAAAGTC AGGCAGTGGT TGAAGAAAGT  
 1101 GTCCGCGCCG GAATACCTGT TCGACAAAGG TGTGCTGGCG GCTGCGGCGG  
 1151 CTGTGAGGTT GGACNGCGGC AGGGCGGCTT TGCGGAGAT CGGCAGGTTG  
 1201 CGGAACTTC CCGAACAGCA GGGCGGTAT TTTACGGCAG ACAATTTGTC  
 1251 CAAAATACAG ATGTTGCCCC TGTCGAAGCT GCCCGACAAA CGGGAGGCTT  
 1301 TGAGGGGTTT GGACAAAGATT ATCGAAAAC CGCTGCGCG CAGTAATACA  
 1351 GAGTTACAGG CAGAGGCATT GGTACAGCGG TCAGTTGTTT ACGATCGGCT  
 1401 TGGCAGCGG AAAAAATGA TTTCAGATCT TGAAGGGCG TTAGGCTTG  
 1451 CACCCGATAA CGCTCAGATT ATGAATAATC TGGGCTACAG CCTGCTTTCC  
 1501 GATTCCAAAC GTTTGACGCA AGGCTTCGCC CTGCTTCAGA CGGCATACCA  
 1551 AATCAACCCG GACGATACCG CTGTCAACGA CAGCATAGGC TGGGCGTATT  
 1601 ACCTGAAANG CGACGCGGAA AGCGCGCTGC CGTATCTGCG GTATTCGTTT  
 1651 GAAACGACC CCGAGCCCGA AGTTGCCGCC CATTGGGCG AAGTGTGTTG  
 1701 GGCATTGGGC GAACGCGATC AGGCGGTTGA CGTATGGACG CAGGCGGCAC  
 1751 ACCTACGGG AGACAAGAAA ATATGGCGGG AAACGCTCAA ACGTCACGGC  
 1801 ATCGCATTGC CCCAACCTTC CCGAAAACCT CGGAAATAA

This encodes a protein having amino acid sequence <SEQ ID 44>:

1 MLPARFTILS VLAAALLAGQ AYAAGAADAK PPKEVGKVF R KQORYSEEEI  
 51 KNERARLA AV GERVNQIFTL LGXETALQK QAGTALATYM LMLERTKSPE  
 101 VAERALEMAV SLNAFEQAEM IYQKWRQIEP IPGKAQKRAG WLRNVLRERG  
 151 NQHL DGLEEX LAQADEXQNR RVFLLLAQAA VQDGLAQKA SKAVRRAALR  
 201 YEHLPEAAVA DVVFSVQXRE KEKAIGALQR LAKL DTEILP PTLMTLRLTA  
 251 RYPEILDGF FEQTDTONLS AVWQEMEIMN LVSLHRLDDA YARLNVLLER  
 301 NPNADLYIQA AILANRKEK ASVIDGYAEK AYGRGTGEQR GRAAMTAAMI  
 351 YADRDRYTKV RQWLKKVSAP EYLFDKGVLA AAAA VELDXG RAALRQIGRV  
 401 RKLPEQQGRY FTADNL SKIQ MFALSKLPDK REALRGLDKI IEKPPAGSNT  
 451 ELQAEALVQR SVVYDR LGKR KKMISDLERA FRLAPDNAQI MNNLGYSLLS  
 501 DSKRLDEGFA LLQAYQINP DDTAVNDSIG WAYY LKXDAE SALPYLRYSF  
 551 ENDPEPEVAA HLEVLWALG ERDQAVDVWT QAAHLTGDKK IWRETLKRHG  
 601 IALPQPSRKP RK\*

ORF9a and ORF9-1 show 95.3% identity in 614 aa overlap:

	10	20	30	40	50
orf9a.pep	MLPARFTILSVLAAALLAGQAYAG--AADAKPPKEVGKVF RKQORYSEEEIKNERARLA				
orf9-1	MLPNRFKMLTVLATLIAGQVSAAGGAGDMKQPKPEVGKVF RKQORYSEEEIKNERARLA				
	10	20	30	40	50
orf9a.pep	AVGERVNQIFTL LGXETALQKQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA				
orf9-1	AVGERVNQIFTL LGGETALQKQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA				
	70	80	90	100	110
orf9a.pep	EMIQKWRQIEPIPGKAQKRAGWLRNVLRERGNQHL DGLEEXLAQADEXQNR RVFLLLAQ				
orf9-1	EMIQKWRQIEPIPGKAQKRAGWLRNVLRERGNQHL DGLEEXLAQADEXQNR RVFLLLAQ				

	orf9-1	EMIQKWRQIEPIPGKAQKRAEWLRNVLRERGNQHLDGLEEVLAQADEGNRRVFLLLAQ	130	140	150	160	170	180
5	orf9a.pep	180 190 200 210 220 230 AAVQQDGLAQKASKAVRRAALRYEHLPEAAVADVFSVQXREKEKAIGALQRLAKLDTEI						
	orf9-1	180 190 200 210 220 230 240 AAVQQDGLAQKASKAVRRAALRYEHLPEAAVADVFSVQGREKEKAIGALQRLAKLDTEI						
10	orf9a.pep	240 250 260 270 280 290 LPPTLMTLRLTARKYPEILDGFFEQTDTQNLSAVWQEMEIMNLVSLHRLDDAYARLNVLL						
	orf9-1	240 250 260 270 280 290 300 LPPTLMTLRLTARKYPEILDGFFEQTDTQNLSAVWQEMEIMNLVSLHRLDDAYARLNVLL						
15	orf9a.pep	300 310 320 330 340 350 ERNPNADLYIQAAILAANRKEKXASVIDGYAEKAYGRGTGEQRGRAAMTAAMIYADRRDYT						
20	orf9-1	300 310 320 330 340 350 360 ERNPNADLYIQAAILAANRKEKXASVIDGYAEKAYGRGTGEQRGRAAMTAAMMYADRRDYA						
25	orf9a.pep	360 370 380 390 400 410 KVRQWLKKVSAPEYLFDKGVLAAAAVELDXGRAALRQIGRVKRLPEQQGRYFTADNLSK						
	orf9-1	360 370 380 390 400 410 420 KVRQWLKKVSAPEYLFDKGVLAAAAVELDGGRAALRQIGRVKRLPEQQGRYFTADNLSK						
30	orf9a.pep	420 430 440 450 460 470 IQMFALSKLPDKREALRGLDKIIEKPPAGSNTLQAEALVQRSVVYDRLGKRKKMISDLE						
	orf9-1	420 430 440 450 460 470 480 IQMLALSKLPDKREALRGLDKIIEKPPAGSNTLQAEALVQRSVVYDRLGKRKKMISDLE						
35	orf9a.pep	480 490 500 510 520 530 RAFRLAPDNAQIMNMLGYSLLSDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLKXD						
	orf9-1	480 490 500 510 520 530 540 RAFRLAPDNAQIMNMLGYSLLSDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLKGD						
40	orf9a.pep	540 550 560 570 580 590 AESALPYLRYSFENDPEPEVAHLGEVLWALGERDQAVDVWTQAAHLTGDKKIWRETLLR						
	orf9-1	540 550 560 570 580 590 600 AESALPYLRYSFENDPEPEVAHLGEVLWALGERDQAVDVWTQAAHLTGDKKIWRETLLR						
45	orf9a.pep	600 610 HGIALPQPSRKPRKX						
50	orf9-1	600 610 HGIALPQPSRKPRKX						

Homology with a predicted ORF from *N.gonorrhoeae*

ORF9 shows 82.8% identity over a 163aa overlap with a predicted ORF (ORF9.ng) from *N. gonorrhoeae*:

55	Orf9	RFKMLTVLTATLIAGQVSAAGGGAGDMKQPKEVGKVFRRKQQRYSSEEEIKNERAR	54
	orf9ng	MIMLPARFTILSVLAAALLAGQAYAA--GAADVLPKEVGKVLKRRYSEEEIKNERAR	58
60	orf9	LAAVGERVNQIFTLGGETALQKGQAGTALATYMLMLERTKSP EVAERALEMAVSLNAFE	114
	orf9ng	LAAVGERVNRVFTLLGGETALQKGQAGTALATYMLMLERTKSP EVAERALEMAVSLNAFE	118
65	orf9	QAEMIQKWRQIEPIPGKAQKRAEWLRNVLRERGNQHLDGREEVLAQADEGQ	166
	orf9ng	QAEMIQKWRQIEPIPGKAQKRAEWLRNVLRERGNQHLDGREEVLAQADEGQ	178

The ORF9ng nucleotide sequence <SEQ ID 45> was predicted to encode a protein having including acid sequence <SEQ ID 46>:

```

1  MIMLPARFTI LSVLAAALLA GQAYAAGAAD VELPKEVGKV LRKHRRYSEE
51 EIKNERARLA AVGERVNRVF TLLGGETALQ KGQAGTALAT YMLMLERTKS
101 PEVAERALEM AVSLNAFEQA EMIYQKWRQI EPIPGEAQKP AGWLRNVLKE
151 GGNPHLDRLE EVPAQSDYVH QPMIFLLLVQ AAVQHGGVAQ KPSKAVRPAA
201 YNYEVLPETA GADAVFCVQG PQYEKAIQSF PFCGRNPQTE NIAPPFNELF
251 RPTARPISPK LLQRFRTPEP NLAKPFRPPG PEMETYQTGF PRPLTRNNPT

```

Amino acids 1-28 are a putative leader sequence, and 173-189 are predicted to be a transmembrane domain.

Further sequence analysis revealed the complete length ORF9ng DNA sequence <SEQ ID 47>:

```

1  ATGTTACCCG CCCGTTTCAC TATTTTATCT GTCCTCGCAG CAGCCCTGCT
51  TGCCGGACAG GCGTATGCTG CCGGCGCGGC GGATGTGGAG CTGCCGAAGG
101 AAGTCGGAAA GGTTTTAAAG AAACATCGGC GTTACAGCGA GGAAGAAATC
151 AAAAACGAAC CGCACGGGCT TCGGCAGTG GGCGAACGGG TCAACAGGGT
201 GTTTACGCTG TTGGGCGGTG AAACGGCTTT GCAGAAAGGG CAGGCGGGAA
251 CGGCTCTGGC AACCTATATG CTGATGTTGG AACGCACAAA ATCCCCCGAA
301 GTCGCCGAAC GCGCCTTGGA AATGGCCGTG TCGTGAACG CGTTTGAACA
351 GGCGGAAATG ATTTATCAGA AATGgcggca gatcgagcct ataCcgsgtg
401 aggcgcaaaa accgCcgggG tggctgcgga acgtattgaa ggaagggGGa
451 aaTCAGCATC TGGACgggtt gaaagaggTG CtggcgcaAT cggacgatGT
501 GCAAAAacgc aggaTATTTT TGCTGCTGGT GCAAGCCGCC GTGCagcagg
551 gTGGGTGGC TCAAAAAGCA TCGAAAGCGG TTCGcgtgc GGcgttgaAG
601 TATGAACATC TGCCcgaagc ggcggTTGCC GATGcggTGT TCGGCGTACA
651 GGGACGCGAA AAGGAAAagg caaTCGAAGC TTTGCAGCGT TTGGCGAAGC
701 TCGATACGGA AATATTGCC CCCACTTTAA TGACGTTGCG TCTGACTGCA
751 CGCAAATATC CCGAAATACT CGACGGCTTT TTCGAGCAGA CAGACACCCA
801 AAACCTTTCG GCCGTCTGGC AGGAAATGGA AATTATGAAT CTGGTTTCCC
851 TGCGTAAGCC GGATGATGCC TATGCGCGTT TGAACGTGCT GTTGGAAACAC
901 AACCAGAATG CAAACCTGTA TATTCAGGCG GCGATATTGG CGGCAACCG
951 AAAAGAAGGT GCGTCCGTTA TCGACGGCTA CGCCGAAAAG GCATACGGCA
1001 GGGGACGGG GGAACAGCGG GGCagggcgg caATgacggc GGCATGATA
1051 TATGCCGACC GCAGGGATTA CGCCAAAGTC AGGCAGTGGT TGAAAAAGT
1101 GTCCGCGCCG GAATACCTGT TCGACAAAGG CGTGTGGCG GCTGCGCGCG
1151 CTGCCGAATT GGACGGAGGC CGGGCGGCTT TGCGGCAGAT CGGCAGGGTG
1201 CGGAAACTTC CCGAACAGCA GGGCGGTAT TTTACGGCAG ACAATTGTGTC
1251 CAAAATACAG ATGCTCGCCC TGTCGAAGCT GCCCGACAAA CGGGAAGCCC
1301 TGATCGGGCT GAACAAATC ATCGCCAAAC TTTCGGCGGC GGAAGCAGC
1351 GAACCTTTGG CGGAAGCATT GGCACAGCGT TCCATTATTT ACGaacAGTT
1401 cggCAAACGG GGAATAATGA TTGCCGACCT tgaAACcgcg CTCAAACTTA
1451 CGCCGATAA TGCACAAATT ATGAATAATC TGGGCTACAG CCTGCTTTC
1501 GATTCCAAAC GTTTGGACGA GGGTTTCGCC CTGCTTCAGA CGGCATACCA
1551 AATCAACCCG GACGATACCG CCGTTAACGA CAGCATAGGC TGGGCGTATT
1601 ACCTGAAAGG CGACgcggaA AGCGCGCTGC CGTATCTGcg gtattcgttt
1651 gAAAACGACC CCGAGCCCGA AGTTGCCGCC CATTGGGCG AAGTGTGTG
1701 GGCATTGGGC GAACGCGATC AGGCGGTTGA CGTATGGACG CAGGCGGCAC
1751 ACCTTAGGGG AGACAAGAAA ATATGGCGGG AGACGCTCAA ACGCTACGGA
1801 ATCGCCTTGC CCGAGCCTTC CCGAAAACCC CGGAAATAA

```

This encodes a protein having amino acid sequence <SEQ ID 48>:

```

1  MLPARFTILS VLAAALLAGQ AYAAGAADVE LPKEVGKVLK KHRRYSEEEI
51 KNERARLAAV GERVNRVFTL LGGETALQKG QAGTALATYM LMLERTKSPE
101 VAERALEMAV SLNAFEQAEM IYQKWRQIEP IPGEAQKPAG WLRNVLKEGG
151 NQHLDGLKEV LAQSDDDVQKR RIFLLLVQAA VQQGGVAQKA SKAVRRAALK
201 YEHLPEAAVA DAVFGVQGRE KEKAIEALQR LAKLDTEILP PTLMTLRLTA
251 RKYPEILDGF FEQTDTONLS AVWQEMEIMN LVSLRKPDDA YARLNVLEH
301 NPNANLYIQA AILANRKEG ASVIDGYAEK AYGRGTGEQR GRAAMTAAMI
351 YADRRDYAKV RQWLKKVSAP EYLFDKGVLA AAAAELDGG RAALRQIGRV
401 RKLPEQQGRY FTADNLSKIQ MLALSCLPKD REALIGLNNI IAKLSAAGST
451 EPLAEALAQR SIIYEQFGKR GKMIADLETA LKLPDANAQI MNNLGYSLLS

```



600

610

In addition, ORF9ng shows significant homology with a hypothetical protein from *P.aeruginosa*:

```

5  sp|P42810|YHE3_PSEAE HYPOTHETICAL 64.8 KD PROTEIN IN HEMM-HEMA INTERGENIC REGION
   (ORF3)
   >gi|1072999|pir||S49376 hypothetical protein 3 - Pseudomonas aeruginosa >gi|557259
   (X82071) orf3 [Pseudomonas aeruginosa] Length = 576
   Score = 128 bits (318), Expect = 1e-28
   Identities = 138/587 (23%), Positives = 228/587 (38%), Gaps = 125/587 (21%)

10 Query: 67 VFTLLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQAEMIYQKWR 126
   +++LL E A Q+ + AL+ Y++ ++T+ P V+ERA +A L A ++A W
   Sbjct: 53 LYSLVAELAGQRNRFDIALSNYVVQAQKTRDPGVSERAFRIAEYLGADQEALDTSLLWA 112

15 Query: 127 QIEPIPGEAQKPAG-----WLRNVLKEGGNQHL DGLKEVLAQSDDVQKRRI 172
   + P +AQ+ A ++ VL G+ H D L A++D + +
   Sbjct: 113 RSAPDNLDAQRAAAIQLARAGRYEESMVYMEKVLNGQGDTHFDLALSAAETDPDTRAGL 172

20 Query: 173 FXXXXXXXXXXXXXXXXKASKAVRRAALKYEHLPAAVADAVFGVQGREKEKAIEALQRLA 232
   ++ KY + + A+ Q ++A+ L+ +
   Sbjct: 173 L-----QSF D H L L K K Y P N N G Q L L F G K A L L L Q Q D G R P D E A L T L L E D N S 214

25 Query: 233 KLDTEILPPTLMTLRLTARK-----YPEILDGFFEQTDTQNL SAVVQEMEIMNLVSLRKP 287
   E+ P L + L + K P + G E D + + + + LV +
   Sbjct: 215 ASRHEVAPLLLR SRL L Q S M K R S D E A L P L L K A G I K E H P D D K R V R L A Y A R L ----LVEQNRL 270

30 Query: 288 DDAYARLNVLL EHNPN-----ANLYIQAAI----- 312
   DDA A L++ P+ A +Y++ +
   Sbjct: 271 DDAKAEFAGLVQQFPDDDDDLRFS LALVCLEAQAWDEARIYLEELVERD SHVDAAHFNLG 330

35 Query: 313 -LAANRKEGASVIDGYAEKAYGRGTGEQRGRAAMTAAMIYADRRDYAKVRQWLKKVSAPE 371
   LA +K+ A +D YA+ G G + T ++ A R D A R + P+
   Sbjct: 331 RLAEQKQD TARALDEY AQ--VGP G N D F L P A Q L R Q T D V L L K A G R V D E A A Q R L D K A R S E Q P D 388

40 Query: 372 YLFDKXXXXXXXXXXXXXXXXX RQIGRVRKLPEQQGRYFTADNLSKIQMLALS KLPDKR 431
   Y A L I+ ALS +
   Sbjct: 389 Y-----AIQLYLIEAEALS NNDQQE 408

45 Query: 432 EALIGLNNIIAKLSAAGSTEPLAEALAQRSIIYEQFGKRGKMIADLETALKLTPDNAQIM 491
   +A + + + E L L RS++ E+ +M DL + PDNA +
   Sbjct: 409 KAWQAIQEGLKQYP-----EDL-NLLYTR SMLAEKRNDLAQMEKDLRFVIAREPDNAMAL 462

50 Query: 492 NNLGYSLSDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLKGD AESALPYLRY SFE 551
   N LGY+L + R E L+ A+++NPDD A+ DS+GW Y +G A YLR + +
   Sbjct: 463 NALGYTLADRTTRYGEARELILKAHKLNPDDPAILD SMGWIN YRQGLADAERYLRQALQ 522

55 Query: 552 NDPEPEVA AHLGEVLWALGERDQAVDVWTQAAHLRGDKKIWR ETLKR 598
   P+ EVA AHLGEVLWA G + A +W + + D + R T+KR
   Sbjct: 523 RYPDHEVA AHLGEVLWAQGRQGDARAIWREYLDKQPDSDVLRRTIKR 569

60 Query: 598 gi|2983399 (AE000710) hypothetical protein [Aquifex aeolicus] Length = 545
   Score = 81.5 bits (198), Expect = 1e-14
   Identities = 61/198 (30%), Positives = 98/198 (48%), Gaps = 19/198 (9%)

65 Query: 408 GRYFTADNL-SKIQMLALS KLPDKREALIGLNNIIAKLSAAGSTEPLAEALQ----- 459
   G Y A L K ++LA PDK+E L + +K + + L +
   Sbjct: 335 GNYEDAKRLIEKAKVLA----PDKKEILFLEADYYSKTKQYDKALEILKKLEKDYPNDSR 390

   Query: 460 ----RSIIYEQFGKRGKMIADLETALKLTPDNAQIMNNGYSLLS--DSKRLDEGFALLQ 513
   +I+Y+ G L A++L P+N N LGYSL L +R++E L++
   Sbjct: 391 VYFMEAIVYDNLGDIKNAEKALRKAIELDPENPDYNYLGYSL LLYGKERVEEAELIK 450

   Query: 514 TAYQINPDDTAVNDSIGWAYYLKGD AESALPYLRY SF-ENDPEPEVA AHLGEVLWALGER 572
   A + +P++ A DS+GW YYLKG D E A+ YL + E +P V H+G+VL +G+
   Sbjct: 451 KALEKDPENPAYIDSMGWVYYLKG DYERAMQYLLKALREAYDDPVVNEHVG DVLLKMGYK 510

   Query: 573 DQAVDVWTQAAHLRGDKK 590
   ++A + + +A L + K
   Sbjct: 511 EARNYYERALKLLEEGK 528

```

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 7

5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 49>:

```

1 AACCTCTACG CCGGCCCGCA GACCACATCC GTCATCGCAA ACATCGCCGA
51 CAACCTGCAA CTGGCCAAAG ACTACGGCAA AGTACACTGG TTCGCCTCCC
101 CGCTCTTCTG GCTCCTGAAC CAACTGCACA ACATCATCGG CAACTGGGGC
151 TGGGCGATTA TCGTTTAAAC CATCATCGTC AAAGCCGTAC TGTATCCATT
201 GACCAACGCC TCTTACCGCT CTATGGCGAA AATGCGTGCC GCCGCACCCA
251 AACTGCAAGC CATCAAAGAG AAATACGGCG ACGACCGTAT GGCGCAACAA
301 CAGGCGATGA TGCAGCTTTA CACAGACGAG AAAATCAACC CGCTGGGCG
351 GCTGCCTGCC TATGCTGTTG CAAATCCCGG TCTTCATCGG ATTGTATTGG
401 GCATTGTTCT CCTCCGTAGA ATTGCGCCAG GCACCTTGGC TGGGTGGGAT
15 451 TACCGACCTC AGCCGCGCCG ACCCTACTA CATCCTGCC ATCATTATGG
501 CGGCAACGAT GTTCGCCCAA ACTTATCTGA ACCCGCCGCC GACCGACCCG
551 ATGCAGGCGA AAATGATGAA AATCATGCCG TTGGTTTTCT CsgwCtTGTT
601 CTTCTTCTTC CCTGCCGgks TGGTATTGTA CTGGGTAGTC AACACCTCC
651 TGACCATCGC CCAGCAATGG CACATCAACC GCAGCATCGA AAAACACGC
20 701 GCCCAAGGCG AAGTCGTTTC CTAA

```

This corresponds to the amino acid sequence <SEQ ID 50; ORF11>:

```

1 . . NLYAGPQTTS VIANIADNLQ LAKDYGKVHW FASPLFWLLN QLHNIIGNWG
51 WAIIVLTIIV KAVLYPLTNA SYRSMAMRA AAPKLQAIKE KYGDDRMAQQ
101 QAMQLYTDE KINPLGGCLP MLLQIPVFIG LYWALFASVE LRQAPWLGI
25 151 TDLSRADPYI ILPIIMAATM FAQTYLNPPP TDFMQAKMMK IMPLVFSXXF
201 FFFPAGXVLY WVVNNLLTIA QQWHINRSIE KQRAQGEVVS *

```

Further sequence analysis revealed the complete DNA sequence <SEQ ID 51>:

```

1 ATGGATTTTA AAAGACTCAC GCGGTTTTTC GCCATCGCGC TGGTGATTAT
51 GATCGGCTGG GAAAAGATGT TCCCCACTCC GAAGCCAGTC CCCGCGCCCC
101 AACAGGCAGC ACAACAACAG GCGGTAACCG CTTCGCGCGA AGCCGCGCTC
151 GCGCCCGCAA CGCCGATTAC CGTAACGACC GACACGGTTC AAGCCGTCAT
201 TGATGAAAAA AGCCGCGACC TCGCGCGGCT GACCTGCTC AAATACAAAG
251 CAACCGGCGA CGAAAATAAA CCGTTCATCC TGTTTGCGCA CGGCAAAGAA
301 TACACCTACG TCGCCCAATC CGAACTTTTG GACGCGCAGG GCAACAACAT
35 351 TCTAAAAGGC ATCGGCTTTA GCGCACCAGG AAAACAGTAC AGCTTGAAG
401 GCGACAAAGT TGAAGTCCGC CTGAGCGCGC CTGAAACACG CGGTCTGAAA
451 ATGACAAAG TTTATACTTT CACCAAAGGC AGCTATCTGG TCAACGTCCG
501 CTTCGACATC GCCAACGGCA GCGGTCAAAC CGCCAACCTG AGCGCGGACT
551 ACCGCATCGT CCGGACCCAC AGCGAACCAG AGGGTCAAGG TTACTTTACC
40 601 CACTCTTACG TCGGCCCTGT TGTTTATACC CCTGAAGGCA ACTTCCAAAA
651 AGTCAGCTTT TCCGACTTGG ACGACGATGC CAAATCCGGC AAATCCGAGG
701 CCGAATACAT CCGCAAAACC CCGACCGGCT GGCTCGGCAT GATTGAACAC
751 CACTTCATGT CCACCTGGAT TCTCCAACCT AAAGGCAGAC AAAGCGTTTG
801 CGCCGCGAGC GAGTGCAACA TCGACATCAA ACGCCGCAAC GACAAGCTGT
45 851 ACAGCACCAG CGTCAGCGTG CCTTTAGCCG CCATCCAAAA CGGCGCGAAA
901 GCGGAAGCCT CCATCAACCT CTACGCCGGC CCGCAGACCA CATCCGTCAT
951 GCCAAACATC GCCGACAACC TGCAACTGGC CAAAGACTAC GGCAAGTAC
1001 ACTGGTTTCG CTCCCGCTC TTCTGGCTCC TGAACCAACT GCACAACATC
50 1051 ATCGGCAACT GGGGCTGGGC GATTATCGTT TTAACCATCA TCGTCAAAGC
1101 CGTACTGTAT CCATTGACCA ACGCTCTTA CCGCTCTATG GCGAAATGC
1151 TGCCCGCCGC ACCCAACTG CAAGCCATCA AAGAGAAATA CGGCGACGAC
1201 CGTATGGCGC AACAACAGGC GATGATGCAG CTTTACACAG ACGAGAAAT
1251 CAACCCGCTG GCGCGCTGCC TGCCATATGT GTTGCAATC CCCGTCTTCA
55 1301 TCGGATTGTA TTGGGCATTG TTCGCCTCCG TAGAATTGCG CCAGGCACCT
1351 TGGCTGGGTT GGATTACCGA CCTCAGCCGC GCCGACCCCT ACTACATCCT
1401 GCCCATCATT ATGGCGGCAA CGATGTTTCG CCAAACCTAT CTGAACCCGC
1451 CGCCGACCGA CCCGATGCAG GCGAAAATGA TGAAAATCAT GCCGTGTTG
1501 TTCTCCGTCA TGTCTTCTT CTCCCTGCC GGTCTGGTAT TGTACTGGGT

```

1551 AGTCAACAAC CTCCTGACCA TCGCCCAGCA ATGGCACATC AACCGCAGCA  
1601 TCGAAAAACA ACGCGCCCAA GGCGAAGTCG TTTCCTAA

This corresponds to the amino acid sequence <SEQ ID 52; ORF11-1>:

5	1	MDFKRLTAFF	AIALVMIGW	EKMFPPTPKPV	PAPQQAQQQ	AVTASAEAAAL
	51	APATPITVTT	DTVQAVIDEK	SGDLRRLTLL	KYKATGDENK	FFILFGDGKE
	101	YTYVAQSELL	DAQGNNILKG	IGFSAPKKQY	SLEGDKVEVR	LSAPETRGLK
	151	IDKVYTFTKG	SYLVNVRFDI	ANGSGQTANL	SADYRIVRDH	SEPEGQGYFT
	201	HSYVGPVVYT	PEGNFQKVSF	SDLDDDAKSG	KDEALYSTSR	PTGWLGMIEH
10	251	HFMSTWLQPL	KGRQSVCAAG	ECNIDIKRRN	DKLYSTSVSV	PLAAIQNGAK
	301	AEASINLYAG	PQTTSVIANI	ADNLQLAKDY	GKVHWFASPL	FWLLNQLHNI
	351	IGNWGWAIIV	LTIIVKAVLY	PLTNASYRSM	AKMRAAAPKL	QAIKEYGDD
	401	RMAQQQAMMQ	LYTDEKINPL	GGCLPMLLQI	PVFIGLYWAL	FASVELRQAP
	451	WLGWITDLSR	ADPYIYLPII	MAATMFAQTY	LNPPPTDPMQ	AKMMKIMPLV
	501	FVSMFFFFPA	GLVLYWVVNN	LLTIAQQWHI	NRSIEKQRAQ	GEVVS*

15 Computer analysis of this amino acid sequence gave the following results:

Homology with a 60kDa inner-membrane protein (accession P25754) of *Pseudomonas putida*

ORF11 and the 60kDa protein show 58% aa identity in 229 aa overlap (BLASTp).

20	ORF11	2	LYAGPQTTSVIANIADNLQAKDYGKVVHFASPLFWLLNQLHNIIGNWGWAIIVLTIIVK	61
	60K	324	LYAGP+ S + ++ L+L DYG + + A P+FWLL +H+++GNWGW+IIVLT+++K LYAGPKIQSKLKELSPGLELTVDYGFLWFIAQPIFWLLQHIHSLGNWGSIIIVLTMLIK	383
25	ORF11	62	AVLYPLTNASYSRMAKMRAAAPKLQAIKEKYGDDRXXXXXXXXXXLYTDEKINPLGGCLPM	121
	60K	384	+ +PL+ ASYSRMA+MRA APKL A+KE++GDDR LY EKNINPLGGCLP+ GLFFPLSAASYSRSMARMRAVAPKLAALKERFGDDRQKMSQAMMELYKKEKINPLGGCLPI	443
30	ORF11	122	LLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYIILPIIMAATMFAQTYLNPPPT	181
	60K	444	L+Q+PVF+ LYW L SVE+RQAPW+ WITDLS DP++ILPIIM ATMF Q LNP P LVQMPVFLALYWVLLESVEMRQAPWILWITDLSIKDFFFILPIIMGATMFIQQRNLNPTTP	503
30	ORF11	182	DPMQAKMMKIMPLVXXXXXXXXXPAGXVLYWVVNNLLTIAQQWHINRSIE	230
	60K	504	DPMQAK+MK+MP++ PAG VLYWVVNN L+I+QQW+I R IE DPMQAKVMKMMPIIFTFFFLWFPAGVLYWVVNNCLSSISQOWYITRIE	552

Homology with a predicted ORF from *N.meningitidis* (strain A)

35 ORF11 shows 97.9% identity over a 240aa overlap with an ORF (ORF11a) from strain A of *N. meningitidis*:

[illegible]



220 230 240  
orfl1.pep WVNNLLTIAQQWHINRSIEKQRAQGEVVSX  
5 orfl1a WVNNLLTIAQQWHINRSIEKQRAQGEVVSX  
520 530 540

The complete length ORF11a nucleotide sequence <SEQ ID 53> is:

1 ANGGATTTTA AAAGACTCAC NNGGTTTTTC GCCATCGCAC TGGTGATTAT  
51 GATCGGATNG NAAANGATGT TCCCCACTCC GAAGCCCGTC CCCGCGCCCC  
10 101 AACAGACGGC ACAACAACAG GCCGTAANCG CTCCGCCGA AGCCGCGCTC  
151 GCGCCCGNAN CGCCGATTAC CGTAACGACC GACACGGTTC AAGCCGTCAT  
201 TGATGAAAAA AGCGGCGACC TCGCGCGGCT GACCTTGCTC AAATACAAAG  
251 CAACCGGCGA CNAATAATAA CCGTTCATCC TGTTTGGCGA CGGCAANAA  
301 TACACCTACN TCGCCANTC CGAACTTTTG GACGCGCAGG GCAACACAT  
15 351 TCTAAAAGGC ATCGGCTTTA GCGCACCAG AAAACAGTAC AGCTTGAAG  
401 GCGACAAAAGT TGAAGTCCGC CTGAGCGCAC CTGAAACACG CGGTCTGAAA  
451 ATCGACAAAG TTTTACTTTT CACCAAGGC AGCTATCTGG TCAACGTCCG  
501 CTTCGACATC GCCAACGGCA GCGGTCAAAC CGCCAACCTG AGCGCGGACT  
551 ACCGCATCGT CCGCGACCAC AGCGAACCAG AGGGTCAAGG CTACTTTACC  
20 601 CACTCTTACG TCGGCCCTGT TGTTTATACC CCTGAAGGCA ACTTCCAAAA  
651 AGTCAGCTTC TCCGACTTGG ACGACGATGC CAANTCCGN AAATCCGAGG  
701 CCGAATACAT CCGCAAAACC CNGACCGGCT GGCTCGGCAT GATTGAACAC  
751 CACTTCATGT CCACCTGGAT CCTCCAACCC AAAGGCGGAC AAAGCGTTT  
801 CGCCGCTGGC GACTGNGTA TNGACATCA ACGCCGCAAC GACAAGCTGT  
25 851 ACAGCACCAG CGTCAGCGTG CCTTTAGCCG CTATCCAAA CGGTGCGAAA  
901 TCCNAAGCCT CCATCAACCT CTACGCCGGC CCACAGACCA CATCNGTTAT  
951 CGCAACATC GCCGACAACC TGCAACTGN CAAAGACTAC GGCAAGTAC  
1001 ACTGGTTTCG CTCCCCCTC TTTTGGCTT TGAACCACT GCACAACATC  
1051 ATCGGCAACT GGGGTGGGC GATTATCGTT TTAACCATCA TCGTCAAAGC  
30 1101 CGTACTGTAT CCATTGACCA ACGCCTCTTA CCGTTCGATG GCGAAAATGC  
1151 GTGCCGCCGC GCCCAAACG CAAGCCATCA AAGAGAAATA CGGCGACGAC  
1201 CGTATGGCGC AGCAACAAGC CATGATGCAG CTTTACACAG ACGAGAAAAT  
1251 CAACCGCTG GCGCGCTGCC TGCCTATGCT GTTGCAAATC CCCGTCTTCA  
1301 TCGGATTGTA TTGGGCATTG TTCGCCTCCG TAGAATTGCG CCAGGCACCT  
35 1351 TGCTGGGTG GATTACCGA CCTCAGCCGC GCCGACCNT ACTACATCCT  
1401 GCCCATCATT ATGGCGGCAA CGATGTTTCG CCAAACCTAT CTGAACCCGC  
1451 CGCCGACCGA CCCGATGCAG GCGAAAATGA TGAAAATCAT GCCTTTGGTT  
1501 NTNTCNNNA NGTCTTCNN CTTCCTGCC GGTCTGGTAT TGTACTGGGT  
1551 GATCAACAAC CTCCTGACCA TCGCCAGCA ATGGCACATC AACCGCAGCA  
40 1601 TCGAAAAACA ACGCGCCAA GCGGAAGTCG TTTCTTAA

This encodes a protein having amino acid sequence <SEQ ID 54>:

1 XDFKRLTXFF AIALVIMIGX XXMFPTPKPV PAPQQTAAQQ AVXASAEAL  
51 APXXPITVTT DTVQAVIDEK SGDRLRLTLL KYKATGDXNK PFILFGDGKX  
45 101 YTYXAXSELL DAQGNILKG IGFSAPKKQY SLEGDKVEVR LSAPETRGLK  
151 IDKVYTFTEG SYLVNVRFDI ANGSGQTANL SADYRIVRDH SEPEGQGYFT  
201 HSYVGPVVTY PEGNFQKVSF SDLDDAXSG KSEAEYIRKT XTGWLMGIEH  
251 HFMSTWILQP KGGQSVCAAG DCXXDIKRRN DKLYSTSVSV PLAAIQNGAK  
301 SXASINLYAG PQTTSVIANI ADNLQLXKDY GKVHWFASPL FWLLNQLHNI  
351 IGNGWGAIIV LTIIKAVLY PLTNASYRSM AKMRAAAPKL QAIKEKYGDD  
50 401 RMAQQQAMMQ LYTDEKINPL GGCLPMLLQI PVFIGLYWAL FASVELRQAP  
451 WLGWITDLR ADPPYILPII MAATMFAQTY LNPPTDPMQ AKMMKIMPLV  
501 XSXXFFXFPA GLVLYWVINN LLTIAQQWHI NRSIEKQRAQ GEVVS\*

ORF11a and ORF11-1 show 95.2% identity in 544 aa overlap:

55 orfl1a.pep 10 20 30 40 50 60  
orfl1-1 10 20 30 40 50 60  
60 orfl1a.pep 70 80 90 100 110 120  
orfl1-1 70 80 90 100 110 120

		70	80	90	100	110	120
5	orf11a.pep	130	140	150	160	170	180
	orf11-1	130	140	150	160	170	180
10	orf11a.pep	190	200	210	220	230	240
	orf11-1	190	200	210	220	230	240
15	orf11a.pep	250	260	270	280	290	300
	orf11-1	250	260	270	280	290	300
20	orf11a.pep	310	320	330	340	350	360
	orf11-1	310	320	330	340	350	360
25	orf11a.pep	370	380	390	400	410	420
	orf11-1	370	380	390	400	410	420
30	orf11a.pep	430	440	450	460	470	480
	orf11-1	430	440	450	460	470	480
35	orf11a.pep	490	500	510	520	530	540
	orf11-1	490	500	510	520	530	540
40	orf11a.pep	550	560	570	580	590	600
	orf11-1	550	560	570	580	590	600
45	orf11a.pep	610	620	630	640	650	660
	orf11-1	610	620	630	640	650	660
50	orf11a.pep	670	680	690	700	710	720
	orf11-1	670	680	690	700	710	720

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF11 shows 96.3% identity over a 240aa overlap with a predicted ORF (ORF11.ng) from *N. gonorrhoeae*:

55	Orf11	NLYAGPQTTSVIANIADNLQAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIVLT	57
	orf11ng	MAVNLYAGPQTTSVIANIADNLQAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIVLT	60
60	orf11	IIVKAVLYPLTNASYRSMAKMRAAAPKLQAIKEKYGDDRMAQQQAMMQLYTDEKINPLGG	117
	orf11ng	IIVKAVLYPLTNASYRSMAKMRAAPELQTIKEKYGDDRMAQQQAMMQLFEDEEINPLGG	120
	orf11	CLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYYILPIIMAATMFAQTYLN	177
65	orf11ng	CLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYYILPIIMAATMFAQTYLN	180

orf11 PPPTDPMQAKMMKIMPLVFSXXFFFFPAGXVLYWVNNLLTIAQQWHINRSIEKQRAQGE 237  
 |||||  
 orf11ng PPPTDPMQAKMMKIMPLVFSVMFFFFPAGLVLYWVNNLLTIAQQWHINRSIEKQRAQGE 240  
 5 orf11 VVS 240  
 |||  
 orf11ng VVS 243

An ORF11ng nucleotide sequence <SEQ ID 55> was predicted to encode a protein having amino acid sequence <SEQ ID 56>:

10 1 MAVNLYAGPQ TTSVIANIAD NLQLAKDYGK VHWFASPLFW LLNQLHNIIG  
 51 NWGWAIVVLT IIVKAVLYPL TNASYRSMK MRAAAPQLQT IKEKYGDDRM  
 101 AQQQAMMQLF EDEEINPLGG CLPMLLQIPV FIGLYWALFA SVELRQAPWL  
 151 GWITDLRAD PYYILPIIMA ATMFAQTYLN PPPTDPMQAK MMKIMPLVFS  
 201 VMFFFFPAGL VLYWVNNLL TIAQQWHINR SIEKQRAQGE VVS\*

15 Further sequence analysis revealed the complete gonococcal DNA sequence <SEQ ID 57> to be:

1 ATGGATTTTA AAAGACTCAC GGC GTTTTTC GCCATCGCGC TGGTGATTAT  
 51 GATCGGCTGG GAAAAAATGT TCCCCACCCC GAAACCCGTC CCCGCGCCCC  
 101 AACAGGCGGC ACAAAAACAG GCAGCAACCG CTTCGCGCGA AGCCGCGCTC  
 151 CGGCCGCGAA CGCCGATTAC CGTAACGACC GACACGGTTC AAGCCGTAT  
 201 TGATGAAAAA AGTGGCGACC TGGCGCGGCT GACCCTGCTC AAATACAAAG  
 251 CAACCGGCGA CGAAAAACAA CCGTTCGTCC TGTTCGCGA CGGCAAGAA  
 301 TACACCTACG TCGCCCAATC CGAACTTTG GACGCGCAGG GCAACAACAT  
 351 TCTGAAAGGC ATCGGCTTTA GCGCACCAGG AAAACAGTAC ACCCTCAACG  
 401 GCGACACAGT CGAAGTCCGC CTGAGCGCGC CGGAAACCAA CGGACTGAAA  
 451 ATCGACAAAG TCTATACCTT TACCAAGAC AGCTATCTGG TCAACGTCCG  
 501 CTTTCGACATC GCGCAACGGA GCGGTCAAAC CGCCAACCTG AGCGCGGACT  
 551 ACCGCATCGT CCGCGACCAC AGCGAACCCG AGGGTCAAGG CTACTTTACC  
 601 CACTCTTACG TCGGCCTGT TGTTCATACC CCTGAAGGCA ACTTCCAAAA  
 651 AGTCAGCTTC TCCgacTTgg acgACGATGC gaaaTccggc aaATccgagg  
 701 ccgaataaca TCGCAAAACC ccgaccggtt ggtcgcgcat gattgaacac  
 751 cacttcatgt ccacctggat cctccAAcct aaaggcggcc aaaacgtttg  
 801 cgcccaggga gactgcccga tcgacattaa aCgcccgaac gacaagctgt  
 851 acagcgcaag cgtcagcgtg cctttaaccg ctatcccaac ccgggggcca  
 901 aaaccgaaaa tggcggTCAA CCTGTATGCC GGTCCGCAAA CCACATCCGT  
 951 TATCGCAAAC ATCGCcgacA ACCTGCAACT GGCAAAAGAC TACGGTAAAG  
 1001 TACACTGGTT CGCATCGCCG CTCTTCTGGC TCCTGAACCA ACTGCACAAC  
 1051 ATTATCGGCA ACTGGGCTG GGCAATCGTC GTTTTGACCA TCATCGTCAA  
 1101 AGCCGTACTG TATCCATTGA CCAACGcctc ctACCGTTCG ATGGCGAAAA  
 1151 TGCGTGccgc cgcacCcaaA CTGCAGACCA TCAAAGAAAA ATAcgGCGAC  
 1201 GACCGTATGG CGCAACAGCA AGCGATGATG CAGCTTTACA AAgacgAGAA  
 1251 AATCAACCCG CTGGGCGGCT GTctgcctat gctgttgCAA ATCCCGTCT  
 1301 TCATCGGCTT GACTGGGCA TTGTTGCCT CCGTAGAATT GCGCCAGGCA  
 1351 CCTTGCTGG GCTGGATTAC CGACCTCAGC CGCGCCGACC CCTACTACAT  
 1401 CCTGCCATC ATTATGGCGG CAACGATGTT CGCCCAAACC TATCTGAACC  
 1451 CGCCGCGGAC CGACCGGATG CAGGCGAAAA TGATGAAAT CATGCCGTTG  
 1501 GTTTTCTCCG TCATGTCTT CTTCTTCCCT GCCGGTTGG TTCTCTACTG  
 1551 GGTGGTCAAC AACCTCTGA CCATCGCCCA GCAGTGGCAC ATCAACCGCA  
 1601 GCATCGAAAA ACAACGCGCC CAAGCGAAG TCGTTTCCTA A

This encodes a protein having amino acid sequence <SEQ ID 58; ORF11ng-1>:

50 1 MDFKRLTAFF AIALVIMIGW EKMFPKPKV PAPQAAQKQ AATASAEAL  
 51 APATPITVTT DTVQAVIDEK SGDLRRLTLL KYKATGDENK PFVLFQDGKE  
 101 YTYVAQSELL DAQGNILKG IGFSAPKKQY TLNGDTEVEV LSAPETNGLK  
 151 IDKVYTFKDI SYLVNVRFDI ANGSGQTANL SADYRIVRDH SEPEGQGYFT  
 201 HSYVGPVVYT PEGNFQKVSF SDLDDDAKSG KSEAERYRKT PTGWLGMIEH  
 55 251 HFMSTWILQP KGGQNVCAQG DCRIDIKRRN DKLYSASVSV PLTAIPTRGP  
 301 KPKMAVNLYA GPQTTSVIAN IADNLQAKD YGKVHWFASP LFLLNQLHN  
 351 IIGNWGWAIV VLTIIKAVL YPLTNASYR MAKMRAAPK LQTIKEYGD  
 401 DRMAQQQAMM QLYKDEKINP LGGCLPMLLQ IPVFIGLYWA LFASVELRQA  
 451 PWLGWITDLS RADPYYILPI IMAATMFAQT YLNPPPTDPM QAKMMKIMPL  
 60 501 VFSVMFFFFP AGLVLYWVNN LLTIAQQWH INRSIEKQRA QGEVVS\*

ORF11ng-1 and ORF11-1 shown 95.1% identity in 546 aa overlap:

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60

```

      10      20      30      40      50      60
orf11ng-1.pep MDFKRLTAFFAIALVIMIGWEKMFPTPKVPVPAQQAAQQAATASAEALAPATPITVTT
orf11-1        MDFKRLTAFFAIALVIMIGWEKMFPTPKVPVPAQQAAQQAATASAEALAPATPITVTT
      10      20      30      40      50      60

      70      80      90     100     110     120
orf11ng-1.pep DTVQAVIDEKSGDLRRLTLLKYKATGDENKPFVLFQDGKEYTYVAQSELLDAQNNILKG
orf11-1        DTVQAVIDEKSGDLRRLTLLKYKATGDENKPFVLFQDGKEYTYVAQSELLDAQNNILKG
      70      80      90     100     110     120

      130     140     150     160     170     180
orf11ng-1.pep IGFSAPKKQYTLNGDTVEVRLSAPETNGLKIDKVYFTKDSYLVNVRFDIANGSGQTANL
orf11-1        IGFSAPKKQYSLGDKVEVRLSAPETRGLKIDKVYFTKGSYLVNVRFDIANGSGQTANL
      130     140     150     160     170     180

      190     200     210     220     230     240
orf11ng-1.pep SADYRIVRDHSEPEGQGYFTHSYVGPVVYTPEGNFQKVSFSDLDDDAKSGKSEAEYIRKT
orf11-1        SADYRIVRDHSEPEGQGYFTHSYVGPVVYTPEGNFQKVSFSDLDDDAKSGKSEAEYIRKT
      190     200     210     220     230     240

      250     260     270     280     290     300
orf11ng-1.pep PTGWLGMIEHHFMSTWILQPKGQNVCAQGDICRIDIKRRNDKLYSASVSVPLTAIPTRGP
orf11-1        PTGWLGMIEHHFMSTWILQPKGRQSVCAAGECNIDIKRRNDKLYSTS SVPLAAIQN-GA
      250     260     270     280     290

      310     320     330     340     350     360
orf11ng-1.pep KPKMAVNLYAGPQTTSVIANIADNLQAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIV
orf11-1        KAEASINLYAGPQTTSVIANIADNLQAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIV
      310     320     330     340     350

      370     380     390     400     410     420
orf11ng-1.pep VLTIIIVKAVLYPLTNASYRSMAKMRAAAPKLQTIKEYGDDRMAQQQAMMQLYKDEKINP
orf11-1        VLTIIIVKAVLYPLTNASYRSMAKMRAAAPKLQAIKEYGDDRMAQQQAMMQLYTDEKINP
      370     380     390     400     410

      430     440     450     460     470     480
orf11ng-1.pep LGGCLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYYILPII MAATMFAQT
orf11-1        LGGCLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYYILPII MAATMFAQT
      430     440     450     460     470

      490     500     510     520     530     540
orf11ng-1.pep YLNPPPTDPMQAKMMKIMPLVFSVMFFFFFFPAGLVLYWVNNLLTIAQQWHINRSIEKQRA
orf11-1        YLNPPPTDPMQAKMMKIMPLVFSVMFFFFFFPAGLVLYWVNNLLTIAQQWHINRSIEKQRA
      490     500     510     520     530

orf11ng-1.pep QGEVVSX
orf11-1        QGEVVSX
      540

```

60 In addition, ORF11ng-1 shows significant homology with an inner-membrane protein from the database (accession number p25754):

65 ID 60IM\_PSEPU STANDARD; PRT; 560 AA.  
AC P25754;  
DT 01-MAY-1992 (REL. 22, CREATED)  
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)  
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
DE 60 KD INNER-MEMBRANE PROTEIN. . . .

[illegible]

Based on this analysis, including the homology to an inner-membrane protein from *P. putida* and the predicted transmembrane domains (seen in both the meningococcal and gonococcal proteins), it

is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 8

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 59>:

```

5       1  ..GCCGTCTTAA TCATCGAATT ATTGACGGGA ACGGTTTATC TTTTGGTTGT
      51  NAGCGCGGCT TTGGCGGGTT CGGGCATTGC TTACGGGCTG ACCGGCAGTA
     101  CGCCTGCCGC CGTCTTGACC GNCGCTCTGC TTTCCGCGCT GGGTATTNG
     151  TTCGTACACG CCAAACCCGCG CGTTAGAAAA GTTGAAACGG ATTCATATCA
     201  GGATTTGGAT GCCGGACAAT ATGTCGAAAT CCTCCGNCAC ACAGGCGGCA
    10   251  ACCGTTACGA AGTT.TTTAT CGCGGTACG. ACTGGCAGGC TCAAAATACG
     301  GGGCAAGAAG AGCTTGAACC AGGAACTCGC GCCCTCATTG TCCGCAAGGA
     351  AGGCAACCTT CTTATTATCA CACACCCTTA A
  
```

This corresponds to the amino acid sequence <SEQ ID 60; ORF13>:

```

15       1  ..AVLIIELLTG TVYLLVVSAA LAGSGIAYGL TGSTPAAVLT XALLSALGIX
      51  FVHAKTAVRK VETDSYQDL D AGQYVEILRH TGGNRYEVXY RGTWQAQNT
     101  GQEELEPGTR ALIVRKEGNL LIITHP*
  
```

Further sequence analysis elaborated the DNA sequence slightly <SEQ ID 61>:

```

20       1  ..GCCGTCTTAA TCATCGAATT ATTGACGGGA ACGGTTTATC TTTTGGTTGT
      51  nAGCGCGGCT TTGGCGGGTT CGGGCATTGC TTACGGGCTG ACCGGCAGTA
     101  CGCCTGCCGC CGTCTTGACC GNCGCTCTGC TTTCCGCGCT GGGTATTNG
     151  TTCGTACACG CCAAACCCGCG CGTTAGAAAA GTTGAAACGG ATTCATATCA
     201  GGATTTGGAT GCCGGACAAT ATGTCGAAAT CCTCCGACAC ACAGGCGGCA
     251  ACCGTTACGA AGTTTTTAT CGCGGTACGc ACTGGCAGGC TCAAAATACG
     301  GGGCAAGAAG AGCTTGAACC AGGAACTCGC GCCCTCATTG TCCGCAAGGA
     351  AGGCAACCTT CTTATTATCA CACACCCTTA A
  
```

This corresponds to the amino acid sequence <SEQ ID 62; ORF13-1>:

```

       1  ..AVLIIELLTG TVYLLVVSAA LAGSGIAYGL TGSTPAAVLT XALLSALGIX
      51  FVHAKTAVRK VETDSYQDL D AGQYVEILRH TGGNRYEVFY RGTWQAQNT
     101  GQEELEPGTR ALIVRKEGNL LIITHP*
  
```

30 Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF13 shows 92.9% identity over a 126aa overlap with an ORF (ORF13a) from strain A of *N.meningitidis*:

```

35       orf13.pep      10      20      30      40      50
      AVLIIELLTGTVYLLVVSAA LAGSGIAYGLTGSTPAAVLT XALLSALGIXF
      orf13a      MTVWFVA VAVLIIELLTGTVYLLVVSAA LAGSGIAYGLTGSTPAAVLT AALLSALGIWF
      10      20      30      40      50      60
40       orf13.pep      60      70      80      90     100     110
      VHAKTAVRKVETDSYQDL DAGQYVEILRH TGGNRYEVXYRGTXWQAQNT GQEELEPGTRA
      orf13a      VHAKTAVGK VETDSYQDL DAGQYAEILRHAGGNRYEVFYRGTHWQAQNT GQEELEPGTRA
      70      80      90     100     110     120
45       orf13.pep      120
      LIVRKEGNLLIITHPX
      orf13a      LIVRKEGNLLI IAKPX
50       130
  
```

The complete length ORF13a nucleotide sequence <SEQ ID 63> is:

```

      1  ATGACTGTAT GGTGTGTTGC CGCTGTTGCC GTCTTAATCA TCGAATTATT
      51  GACGGGAACG GTTTATCTTT TGGTTGTCAG CGCGGCTTTG GCGGGTTCGG
    101  GCATTGCTTA CGGGCTGACC GGCAGCACGC CTGCCGCCGT CTTGACCGCC
    151  GCTCTGCTTT CCGCGCTGGG TATTTGGTTC GTACAGCCCA AAACCGCCGT
    201  GGGAAAAGTT GAAACGGATT CATATCAGGA TTTGGATGCC GGGCAATATG
    251  CCGAAATCCT CCGGCACGCA GCGGCAACC GTTACGAAGT TTTTATCGC
    301  GGTACGCACT GGCAGGCTCA AAATACGGG CAAGAAGAGC TTGAACCGG
    351  AACGCGCGCC CTAATCGTCC GCAAGGAAG CAACCTTCTT ATCATCGCAA
    401  AACCTTAA
  
```

This encodes a protein having amino acid sequence <SEQ ID 64>:

```

      1  MTWVFVAAVA VLIIELLTGT VYLLVVSAAAL AGSGIAYGLT GSTPAAVLTA
     51  ALLSALGIWF VHAKTAVGKV ETDSYQDLDA GQYAEILRHA GGNRYEVFVR
    101  GTHWQAQNTG QEELEPGTRA LIVRKEGNLL IIAKP*
  
```

ORF13a and ORF13-1 show 94.4% identity in 126 aa overlap

```

                                10      20      30      40      50      60
orff13a.pep  MTWVFVAAVAVLIIELLTGTVYLLVVSAAALAGSGIAYGLTGSTPAAVLTAALLSALGIWF
orff13-1      AVLIIELLTGTVYLLVVSAAALAGSGIAYGLTGSTPAAVLTXALLSALGIXF
                                10      20      30      40      50

                                70      80      90     100     110     120
orff13a.pep  VHAKTAVRKVETDSYQDLDAQYVEILRHAGGNRYEVFYRGTHWQAQNTGQEELEPGTRA
orff13-1      VHAKTAVRKVETDSYQDLDAQYVEILRHTGGNRYEVFYRGTHWQAQNTGQEELEPGTRA
                                60      70      80      90     100     110

                                130
orff13a.pep  LIVRKEGNLLIIAKPX
orff13-1      LIVRKEGNLLIITHPX
                                120
  
```

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF13 shows 89.7% identity over a 126aa overlap with a predicted ORF (ORF13.ng) from *N.gonorrhoeae*:

```

orff13      AVLIIELLTGTVYLLVVSAAALAGSGIAYGLTGSTPAAVLTXALLSALGIXF  51
orff13ng    MTWVFVAAVAVLIIELLTGTVYLLVVSAAALAGSGIAYGLTGSTPAAVLTAALLSALGIWF  60

orff13      VHAKTAVRKVETDSYQDLDAQYVEILRHTGGNRYEVXYRGTXWQAQNTGQEELEPGTRA 111
orff13ng    VHAKTAVGKVETDSYQDLDTGKYAEILRYTGGNRYEVFYRGTHWQAQNTGQEVFEPGTRA 120

orff13      LIVRKEGNLLIITHP  126
orff13ng    LIVRKEGNLLIIANP  135
  
```

The complete length ORF13ng nucleotide sequence <SEQ ID 65> is:

```

      1  ATGACTGTAT GGTGTGTTGC CGCTGTTGCC GTCTTAATCA TCGAATTATT
     51  GACGGGAACG GTTTATCTTT TGGTTGTCAG CGCGGCTTTG GCGGGTTCGG
    101  GCATTGCTTA CGGGCTGACT GGCAGCACGC CTGCCGCCGT CTTGACCGCC
    151  GCACTGCTTT CCGCGCTGGG CATTTGGTTC GTACATGCCA AAACCGCCGT
    201  GGGAAAAGTT GAAACGGATT CATATCAGGA TTTGGATACC GGAAATATG
    251  CCGAAATCCT CCGATACACA GCGGCAACC GTTACGAAGT TTTTATCGC
    301  GGTACGCACT GGCAGGCACA AAATACGGG CAGGAAGTGT TTGAACCGGG
    351  AACGCGCGCC CTCATCGTCC GCAAAGAAG TAACCTTCTT ATCATCGCAA
    401  ACCCTTAA
  
```

This encodes a protein having amino acid sequence <SEQ ID 66>:

```

1  MTVWFVAAVA VLIIELLTGT VYLLVVSAA AGSGIAYGLT GSTPAAVLTA
51  ALLSALGIWF VHAKTAVGKV ETDSYQDLDT GKYAEILRYT GGNRYEVFYR
101 GTHWQAQNTG QEVFEPGTRA LIVRKEGNLL IIANP*

```

5 ORF13ng shows 91.3% identity in 126 aa overlap with ORF13-1:

```

10 orf13-1.pep          10      20      30      40      50
    AVLIIELLTGTVYLLVVSAAAGSGIAYGLTGSTPAAVLTXALLSALGIXF
    |||||
orf13ng      MTVWFVAAVAVLIIELLTGTVYLLVVSAAAGSGIAYGLTGSTPAAVLTAALLSALGIWF
            10      20      30      40      50      60

15 orf13-1.pep          60      70      80      90     100     110
    VHAKTAVRKVETDSYQDLDTAGQYVEILRHTGGNRYEVFYRGTHWQAQNTGQEELEPGTRA
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf13ng      VHAKTAVGKVETDSYQDLDTGKYAEILRYTGGNRYEVFYRGTHWQAQNTGQEVFEPGTRA
            70      80      90     100     110     120

20 orf13-1.pep          120
    LIVRKEGNLLIITHPX
    |||||
orf13ng      LIVRKEGNLLIIANPX
            130

```

Based on this analysis, including the extensive leader sequence in this protein, it is predicted that ORF13 and ORF13ng are likely to be outer membrane proteins. It is thus predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 9

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 67>:

```

30 1  ATGTwTGATT TCGGTTTGG CGArCTGGTT TTTGTGGCA TTATCGCCCT
    51  GATwGtCCTC GGCCCCGAAC GCsTGCCCGA GGCCGCCCGC AyCGCCGGAC
    101 GGcTCATCGG CAGGCTGCAA CGCTTTGTCTG GcAGCGTCAA ACAGGAATTT
    151 GACACTCAAA TCGAACTGGA AGAACTGAGG AAGGCAAAGC AGGAATTTGA
    201 AGCTGCCGcC GCTCAGGTTC GAGACAGCCT CAAAGAAACC GGTACGGATA
35 251 TGGAAAGCAA TCTGCACGAC ATTTCCGACG GTCTGAAGCC TTGGGAAAAA
    301 CTGCCCCGAAC AGCGGACACC TGCCGATTTC GGTGTCGATG AAAACGGCAA
    351 TCCGCT.TCC CGATGCGGCA AACACCCTAT CAGACGGCAT TTCCGACGTT
    401 ATGCCGTC..

```

This corresponds to the amino acid sequence <SEQ ID 68; ORF2>:

```

40 1  MXDFGLGELV FVGIIALIVL GPERXPEAAR XAGRLIGRLQ RFVGSVKQEF
    51  DTQIELEELR KAKQEFEEAA AQVRDSLKET GTDMEGNLHD ISDGLKPWEK
    101 LPEQRTPADE GVDENGNPXS RCGKHPIRRH FRRYAV..

```

Further work revealed the complete nucleotide sequence <SEQ ID 69>:

```

45 1  ATGTTTGATT TCGGTTTGGG CGAGCTGGTT TTTGTGGCA TTATCGCCCT
    51  GATTGTCTC GGCCCCGAAC GCCTGCCCGA GGCCGCCCGC ACCGCCGGAC
    101 GGCTCATCGG CAGGCTGCAA CGCTTTGTCTG GCAGCGTCAA ACAGGAATTT
    151 GACACTCAAA TCGAACTGGA AGAACTGAGG AAGGCAAAGC AGGAATTTGA
    201 AGTGCCCGCC GCTCAGGTTC GAGACAGCCT CAAAGAAACC GGTACGGATA
50 251 TGGAAAGCAA TCTGCACGAC ATTTCCGACG GTCTGAAGCC TTGGGAAAAA
    301 CTGCCCCGAAC AGCGGACACC TGCCGATTTC GGTGTCGATG AAAACGGCAA
    351 TCCGCTTCCC GATGCGGCAA ACACCCTATC AGACGGCATT TCCGACGTTA
    401 TGCCGTCCGA ACCTTCCTAC GCTTCCGCCG AAACCCTTGG GGACAGCGGG

```



5 451 CAAACCGGCA GTACAGCCGA ACCCGCGGAA ACCGACCAAG ACCGCGCATG  
 501 GCGGGAATAC CTGACTGCTT CTGCCGCCGC ACCCGTCGTA CAGACCGTCG  
 551 AAGTCAGCTA TATCGATACT GCTGTTGAAA CGCCTGTTCC GCACACCACT  
 601 TCCCTGCGCA AACAGGCAAT AAGCCGCAAA CGCGATTTC GTCCGAAACA  
 651 CCGCGCCAAA CCTAAATTGC GCGTCCGTAA ATCATAA

This corresponds to the amino acid sequence <SEQ ID 70; ORF2-1>:

1 MFDFGLGELV FVGIIALIVL GPERLPEAAR TAGRLIGRLQ RFVGSVKQEF  
 51 DTQIELEELR KAKQEFEEAA AQRVDSLKET GTDMEGNLHD ISDGLKPWEK  
 101 LPEQRTPADF GVDENGNPLP DAANTLSDGI SDVMPSESY ASAETLGDSG  
 151 QTGSTAEPAE TDQDRAWREY LTASAAAPVV QTVEVSYIDT AVETPVPHTT  
 201 SLRKQAISRK RDLRPKHSRAK PKLRVRKS\*

Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 71 >:

1 ATGTTTGATT TCGGTTTGGG CGAGCTGGTT TTTGTCGGCA TTATCGCCCT  
 51 GATTGTCTC GGGCCCGAAC GCCTGCCCGA GGCCGCCCGC ACCGCCGGAC  
 151 GGCTCATCG CAGGCTGCAA CGCTTTGTCG GCAGCGTCAA ACAGGAATTT  
 151 GACACGCAAA TCGAACTGGA AGAACTAAGG AAGGCAAAGC AGGAATTTGA  
 201 AGCTGCCGCT GCTCAGGTTT GAGACAGCCT CAAAGAAACC GTACGGATA  
 251 TGGAGGGTAA TCTGCACGAC ATTTCCGACG GTCTGAAGCC TTGGGAAAAA  
 301 CTGCCCGAAC AGCGCACGCC TGCTGATTTC GGTGTCGATG AAAACGGCAA  
 351 TCCCTTTCCC GATGCGGCAA ACACCCTATT AGACGGCATT TCCGACGTTA  
 401 TGCCGTCGTA ACGTTTCTAC GCTTCCGCGC AAACCTTGG GGACAGCGGG  
 451 CAAACCGGCA GTACAGCCGA ACCCGCGGAA ACCGACCAAG ACCGTGCATG  
 501 GCGGGAATAC CTGACTGCTT CTGCCGCCGC ACCCGTCGTA CAGACCGTCG  
 551 AAGTCAGCTA TATCGATACT GCTGTTGAAA CCCTGTTCC GCATACCACT  
 601 TCGCTGCGTA AACAGGCAAT AAGCCGCAAA CGCGATTTC GTCCCTAAATC  
 651 CCGCGCCAAA CCTAAATTGC GCGTCCGTAA ATCATAA

This encodes a protein having amino acid sequence <SEQ ID 72; ORF2a>:

1 MFDFGLGELV FVGIIALIVL GPERLPEAAR TAGRLIGRLQ RFVGSVKQEF  
 51 DTQIELEELR KAKQEFEEAA AQRVDSLKET GTDMEGNLHD ISDGLKPWEK  
 101 LPEQRTPADF GVDENGNPLP DAANTLLDGI SDVMPSESY ASAETLGDSG  
 151 QTGSTAEPAE TDQDRAWREY LTASAAAPVV QTVEVSYIDT AVETPVPHTT  
 201 SLRKQAISRK RDLRPKHSRAK PKLRVRKS\*

The originally-identified partial strain B sequence (ORF2) shows 97.5% identity over a 118aa overlap with ORF2a:

35		10	20	30	40	50	60
	orf2.pep	MXDFGLGELVFVGIIALIVL	GPERXPEAARXAGRLIGRLQ	RFVGSVKQEF	DTQIELEELR		
	orf2a	MFDFGLGELVFVGIIALIVL	GPERLPEAARTAGRLIGRLQ	RFVGSVKQEF	DTQIELEELR		
40		10	20	30	40	50	60
	orf2.pep	KAKQEFEEAA	AQRVDSLKETG	TDMEGNLHD	ISDGLKPWEK	LPEQRTPAD	FGVDENGNPXS
	orf2a	KAKQEFEEAA	AQRVDSLKETG	TDMEGNLHD	ISDGLKPWEK	LPEQRTPAD	FGVDENGNPFP
45		70	80	90	100	110	120
	orf2.pep						
	orf2a						
	orf2.pep	RCGKHPIRRHF	RRYAV				
50		130					
	orf2a	DAANTLLDGI	SDVMPSESY	ASAETLGDSG	QTGSTAEP	AE	TDQDRAWREYLTASAAAPVV
		130	140	150	160	170	180

The complete strain B sequence (ORF2-1) and ORF2a show 98.2% identity in 228 aa overlap:

orf2a.pep MFDFGLGELVFVGIIALIVL GPERLPEAARTAGRLIGRLQ RFVGSVKQEF DTQIELEELR 60  
 118  
 orf2-1 MFDFGLGELVFVGIIALIVL GPERLPEAARTAGRLIGRLQ RFVGSVKQEF DTQIELEELR 60  
 118  
 orf2a.pep KAKQEFEEAA AQRVDSLKETG TDMEGNLHD ISDGLKPWEK LPEQRTPAD FGVDENGNPFP 120  
 118

orf2-1 KAKQFEFAAAAQVRDSLKETGTDMEGNLHDISDGLKPWEKLPEQRTPADFGVDENGNNPLP 120  
 orf2a.pep DAANTLLDGISDVMPSESYASAETLGDSGQTGSTAEPETDQDRAWREYLTASAAAPVV 180  
 5 orf2-1 DAANTLSDGISDVMPSESYASAETLGDSGQTGSTAEPETDQDRAWREYLTASAAAPVV 180  
 orf2a.pep QTVEVSYIDTAVETPVPHTTSLRKQAISRKRDLRPKSRAPKLRVRKSX 229  
 10 orf2-1 QTVEVSYIDTAVETPVPHTTSLRKQAISRKRDLRPKSRAPKLRVRKSX 229

Further work identified a partial DNA sequence <SEQ ID 73> in *N.gonorrhoeae* encoding the following amino acid sequence <SEQ ID 74; ORF2ng>:

1 MFD FGLGELI FVGIIALIVL GPERLPEAAR TAGRLIGRLQ RFVGSVKQEL  
 51 DTQIELEELR KVKQAFEAAA AQVRDSLKET DTDQMNSLHD ISDGLKPWEK  
 15 101 LPEQRTPADF GVDEKGNLSL RYGKHIRRH FFRYAV\*

Further work identified the complete gonococcal gene sequence <SEQ ID 75>:

1 ATGTTTGATT TCGGTTTGGG CGAGCTGATT TTTGTCGGCA TTATCGCCCT  
 51 GATTGTCCTT GGTCCAGAAC GCCTGCCCGA AGCCGCCCGC ACTGCCGGAC  
 101 GGCTTATCGG CAGGCTGCAA CGCTTTGTAG GAAGCGTCAA ACAAGAACTT  
 151 GAACTCAAAA TCGAACTGGA AGAGCTGAGG AAGGTCAAGC AGGCATTCGA  
 201 AGCTGCCGCC GCTCAGGTTT GAGACAGCCT CAAAGAAACC GATACGGATA  
 251 TGCAGAACAG TCTGCACGAC ATTTCCGACG GTCTGAAGCC TTGGGAAAAA  
 301 CTGCCCCAAC AGCGCACGCG tgcgcgatttc gGTGTCGATg AAAacggcaa  
 351 tcccccttccc gATACGGCAA ACACCGTATC AGACGGCATT TCCGACGTTA  
 25 401 TGCCGTCTGA ACGTTCCGAT ACTtccgcCG AAACCCTTGG GGACGACAGG  
 451 CAAACCGGCA GTACAGCCGA ACCTGCGGAA ACCGACAAAG ACCGCGCATG  
 501 GCGGGAATAC CTGactgctt ctgcccgcgc acctgtcgtc Cagagggccg  
 551 tcgaagttag ctaTATCGAT ACTGCTGTTG AAacgcctgt tccgcaCacc  
 601 acttccctgc gcaAACAGGC AATAAACCGC AAACGCGATT TttgtccgaA  
 30 651 ACACCGCGCGc aAACCGAAat tgcgcgtcCG TAAATCATAA

This encodes a protein having the amino acid sequence <SEQ ID 76; ORF2ng-1>:

1 MFD FGLGELI FVGIIALIVL GPERLPEAAR TAGRLIGRLQ RFVGSVKQEL  
 51 DTQIELEELR KVKQAFEAAA AQVRDSLKET DTDQMNSLHD ISDGLKPWEK  
 101 LPEQRTPADF GVDENGNNPL DTANTVSDGI SDVMPSESRD TSAETLGDDR  
 35 151 QTGSTAEPAE TDKDRAWREY LTASAAAPVV QRAVEVSYID TAVETPVPHT  
 201 TSLRKQAINR KRDFCPKHRA KPKLRVRKS\*

The originally-identified partial strain B sequence (ORF2) shows 87.5% identity over a 136aa overlap with ORF2ng:

orf2.pep MXDFGLGELVFVGIIALIVLGPXRPEAARXAGRLIGRLQRFVGSVKQEFDTQIELEELR 60  
 40 orf2ng MFD FGLGELIFVGIIALIVLGPRLPEAARTAGRLIGRLQRFVGSVKQELDTQIELEELR 60  
 orf2.pep KAKQFEFAAAAQVRDSLKETGTDMEGNLHDISDGLKPWEKLPEQRTPADFGVDENGNNPXS 120  
 45 orf2ng KVKQAFEFAAAAQVRDSLKETDTDMQNSLHDISDGLKPWEKLPEQRTPADFGVDEKGNLSL 120  
 orf2.pep RCGKHPIRRHFRRYAV 136  
 orf2ng RYGKHIRRHFRYAV 136

The complete strain B and gonococcal sequences (ORF2-1 & ORF2ng-1) show 91.7% identity in 229 aa overlap:

orf2-1.pep MFD FGLGELVFVGIIALIVLGPRLPEAARTAGRLIGRLQRFVGSVKQEFDTQIELEELR  
 55 orf2ng-1 MFD FGLGELIFVGIIALIVLGPRLPEAARTAGRLIGRLQRFVGSVKQELDTQIELEELR

		10	20	30	40	50	60
		70	80	90	100	110	120
5	orf2-1.pep	KAKQFEAAAAQVRDSLKETGTDMEGNLHDISDGLKPWEKLPEQRTPADFGVDENGNP					
	orf2ng-1	KVKQAFEAAAAQVRDSLKETDTDMQNSLHDISDGLKPWEKLPEQRTPADFGVDENGNP					
		70	80	90	100	110	120
		130	140	150	160	170	180
10	orf2-1.pep	DAANTLSDGISDVMPSESYASAETLGDSGQTGSTAEPQDQDRAWREYLTASAAAPVV					
	orf2ng-1	DTANTVSDGISDVMPSESDTSAETLGDDRQTGSTAEPQDQDRAWREYLTASAAAPVV					
		130	140	150	160	170	180
15		190	200	210	220	229	
	orf2-1.pep	Q-TVEVSYIDTAVETVPVHTTSLRKQAISRKRDFRPHRAKPKLRVRKSX					
	orf2ng-1	QRAVEVSYIDTAVETVPVHTTSLRKQAINRKRDFCPRHRAKPKLRVRKSX					
		190	200	210	220	230	

- 20 Computer analysis of these amino acid sequences indicates a transmembrane region (underlined), and also revealed homology (59% identity) between the gonococcal sequence and the TatB protein of *E.coli*:

25 gnl|PID|e1292181 (AJ005830) TatB protein [Escherichia coli] Length = 171  
 Score = 56.6 bits (134), Expect = 1e-07  
 Identities = 30/88 (34%), Positives = 52/88 (59%), Gaps = 1/88 (1%)

30 Query: 1 MFD FGLGELIFVGIIALIVLGPRLPEAARTAGRLIGRLQRFVGSVKQELDTQIELEELR 60  
 MFD G EL+ V II L+VLGP+RLP A +T I L+ +V+ EL +++L+E +  
 Sbjct: 1 MFDIGFSELLLVFIIIGLVLPQRLPVAVKTVAGWIRALRSLATTVQNELTQELKLQEFQ 60

Query: 61 -KVKQAFEAAAAQVRDSLKETDTDMQNS 87  
 +K+ +A+ + LK + +++ +  
 Sbjct: 61 DSLKKVEKASLTNLTPELKASMDLRQA 88

- 35 Based on this analysis, it was predicted that ORF2, ORF2a and ORF2ng are likely to be membrane proteins and so the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

- 40 ORF2-1 (16kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 3A shows the results of affinity purification of the GST-fusion protein, and Figure 3B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for Western blots (Figure 3C), ELISA (positive result), and FACS analysis (Figure 3D). These experiments confirm that ORF37-1 is a surface-exposed protein, and that it is a useful immunogen.

### Example 10

- 45 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 77>:

1 ATGCAAGCAC GGCTGCTGAT ACCTATTCTT TTTTCAGTTT TTATTTTATC  
 51 CGC.TGCGGG AACTGACAG GTATTCCATC GCATGGCGgA GkTAAACgCT

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60

```

101 TTgCGGTCGA ACAAGAACTT GTGGCCGCTT CTGCCAGAGC TGCCGTTAAA
151 GACATGGATT TACAGGCATT ACACGGACGA AAAGTTGCAT TGTACATTGC
201 CACTATGGGC GACCAAGGTT CAGGcAGTTT GACAGGGGGG TCGCTACTCC
251 ATTGATGCAC kGrTwCsTGG CGAATACATA AACAGCCCTG CCGTCCGTAC
301 CGATTACACC TATCCACGTT ACGAAACCAC CGCTGAAACA ACATCAGGCG
351 GTTTGACAGG TTTAACCCTT TCTTTATCTA CACTTAATGC CCCTGCACTC
401 TCTCGCACC AATCAGACGG TAGCGGAAGT AAAAGCAGTC TGGGCTTAAA
451 TATTGGCGGG ATGGGGGATT ATCGAAATGA AACCTTGACG ACTAACCCGC
501 GCGACACTGC CTTTCTTTCC CACTTGGTAC AGACCGTATT TTTCTGCGC
551 GGCATAGACG TTGTTTCTCC TGCCAATGCC GATACAGATG TGTATTATTA
601 CATCGACGTA TTCGGAACGA TACGCAACAG AACCGAAATG..

```

This corresponds to the amino acid sequence <SEQ ID 78; ORF15>:

15  
20

```

1 MQARLLIPIL FSVFILSACG TLTGIPSHGG XKRFAVEQEL VAASARAAVK
51 DMDLQALHGR KVALYIATMG DQSGSLTGG RYSIDAXXXG EYINSPAVRT
101 DYTYPREYET AETTSGLTGT LTSLSTLNA PALSRQSDG SGSKSSLGLN
151 IGGMGDYRNE TLTTNPRDTA FLSHLVQTVF FLRGIDVVSP ANADTDVFIN
201 IDVFGTIRNR TEM..

```

Further work revealed the complete nucleotide sequence <SEQ ID 79>:

20  
25  
30  
35  
40  
45  
50  
55  
60  
65  
70  
75  
80  
85  
90  
95

```

1 ATGCAAGCAC GGCTGCTGAT ACCTATTCTT TTTTCAGTTT TTATTTTATC
51 CGCCTGCGGG AACTGACAG GTATTCCATC GCATGGCGGA GGTAAACGCT
101 TTGCGGTCGA ACAAGAACTT GTGGCCGCTT CTGCCAGAGC TGCCGTTAAA
151 GACATGGATT TACAGGCATT ACACGGACGA AAAGTTGCAT TGTACATTGC
201 CACTATGGGC GACCAAGGTT CAGGCAGTTT GACAGGGGGT CGCTACTCCA
251 TTGATGCACT GATTCGTGGC GAATACATAA ACAGCCCTGC CGTCCGTACC
301 GATTACACCT ATCCACGTTA CGAAACCACC GCTGAAACAA CATCAGGCGG
351 TTTGACAGGT TTAACCACTT CTTTATCTAC ACTTAATGCC CCTGCACTCT
401 CTCGCACCCA ATCAGACGGT AGCGGAAGTA AAAGCAGTCT GGGCTTAAAT
451 ATTGGCGGGA TGGGGGATTA TCGAAATGAA ACCTTGACGA CTAACCCGCG
501 CGACACTGCC TTTCTTTCCC ACTTGGTACA GACCGTATTT TTCCTGCGCG
551 GCATAGACGT TGTTCCTCCT GCCAATGCCG ATACAGATGT GTTTATTAAAC
601 ATCGACGTAT TCGGAACGAT ACGCAACAGA ACCGAAATGC ACCTATACAA
651 TGCCGAAACA CTGAAAGCCC AAACAAAACCT GGAATATTTC GCAGTAGACA
701 GAACCAATAA AAAATTGCTC ATCAAAACCA AAACCAATGC GTTGAAGCT
751 GCCTATAAAG AAAATTACGC ATTGTGGATG GGGCCGTATA AAGTAAGCAA
801 AGGAATTAAA CCGACGGAAG GATTAATGGT CGATTTCTCC GATATCCGAC
851 CATACGGCAA TCATACGGGT AACTCCGCCC CATCCGTAGA GGCTGATAAC
901 AGTCATGAGG GGTATGGATA CAGCGATGAA GTAGTGCAGC AACATAGACA
951 AGGACAACCT TGA

```

This corresponds to the amino acid sequence <SEQ ID 80; ORF15-1>:

40  
45  
50

```

1 MQARLLIPIL FSVFILSACG TLTGIPSHGG GKRFVEQEL VAASARAAVK
51 DMDLQALHGR KVALYIATMG DQSGSLTGG RYSIDALIRG EYINSPAVRT
101 DYTYPREYET AETTSGLTGT LTSLSTLNA PALSRQSDG SGSKSSLGLN
151 IGGMGDYRNE TLTTNPRDTA FLSHLVQTVF FLRGIDVVSP ANADTDVFIN
201 IDVFGTIRNR TEMHLYNAET LKAQTKLEYF AVDRTNKKLL IKPKTNAFEA
251 AYKENYALWM GPYKVSKEIK PTEGLMVDPS DIRPYGNHTG NSAPSVEADN
301 SHEGYGYSDE VVRQHRQQP *

```

Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 81>:

50  
55  
60

```

1 ATGCAAGCAC GGCTGCTGAT ACCTATTCTT TTTTCAGTTT TTATTTTATC
51 CGCCTGCGGG AACTGACAG GTATTCCATC GCATGGCGGA GGTAAACGCT
101 TTGCGGTCGA ACAAGAACTT GTGGCCGCTT CTGCCAGAGC TGCCGTTAAA
151 GACATGGATT TACAGGCATT ACACGGACGA AAAGTTGCAT TGTACATTGC
201 AACTATGGGC GACCAAGGTT CAGGCAGTTT GACAGGGGGT CGCTACTCCA
251 TTGATGCACT GATTCGTGGC GAATACATAA ACAGCCCTGC CGTCCGTACC
301 GATTACACCT ATCCACGTTA CGAAACCACC GCTGAAACAA CATCAGGCGG
351 TTTGACAGGT TTAACCACTT CTTTATCTAC ACTTAATGCC CCTGCACTCT
401 CGCGCACCCA ATCAGACGGT AGCGGAAGTA AAAGCAGTCT GGGCTTAAAT
451 ATTGGCGGGA TGGGGGATTA TCGAAATGAA ACCTTGACGA CTAACCCGCG
501 CGACACTGCC TTTCTTTCCC ACTTGGTACA GACCGTATTT TTCCTGCGCG
551 GCATAGACGT TGTTCCTCCT GCCAATGCCG ATACGGATGT GTTTATTAAAC
601 ATCGACGTAT TCGGAACGAT ACGCAACAGA ACCGAAATGC ACCTATACAA
651 TGCCGAAACA CTGAAAGCCC AAACAAAACCT GGAATATTTC GCAGTAGACA

```

5  
 701 GAACCAATAA AAAATTGCTC ATCAAACCAA AAACCAATGC GTTTGAAGCT  
 751 GCCTATAAAG AAAATTACGC ATTGTGGATG GGACCGTATA AAGTAAGCAA  
 801 AGGAATTAAA CCGACAGAAG GATTATGGT CGATTCTCC GATATCCAAC  
 851 CATACGGCAA TCATATGGGT AACTCTGCC CATCCGTAGA GGCTGATAAC  
 901 AGTCATGAGG GGTATGGATA CAGCGATGAA GCAGTGCGAC GACATAGACA  
 951 AGGGCAACCT TGA

This encodes a protein having amino acid sequence <SEQ ID 82; ORF15a>:

10  
 1 10  
 51 101  
 151 201  
 251 301  
 MQARLLIPIL FSVFILSACG TLTGIPSHGG GKRFVEQEL VAASARAANK  
 DMDLQALHGR KVALYIATMG DQSGSGLTGG RYSIDALIRG EYINSPAVRT  
 DYTPRYETT AETTSGLTGG LTSLSTLNA PALSRTQSDG SGSKSSLGLN  
 IGGMGDYRNE TLTTNPRDTA FLSHLVQTVF FLRGIDVVP ANADTDVFIN  
 IDVFGTIRNR TEMHLYNAET LKAQTKLEYF AVDRTNKKLL IKPKTNAFEA  
 AYKENYALWM GPYKVSIGIK PTEGLMVDFFS DIQPYGNHMG NSAPSVEADN  
 SHEGYGYSDE AVRRHRQGP \*

15 The originally-identified partial strain B sequence (ORF15) shows 98.1% identity over a 213aa overlap with ORF15a:

20  
 25  
 30  
 35  
 orf15.pep  
 orf15a  
 orf15.pep  
 orf15a  
 orf15.pep  
 orf15a  
 orf15.pep  
 orf15a

40 The complete strain B sequence (ORF15-1) and ORF15a show 98.8% identity in 320 aa overlap:

45  
 50  
 55  
 60  
 orf15a.pep  
 orf15-1  
 orf15a.pep  
 orf15-1  
 orf15a.pep  
 orf15-1  
 orf15a.pep  
 orf15-1

		190	200	210	220	230	240
		250	260	270	280	290	300
5	orf15a.pep	IKPKTNAFEAAAYKENYALWMGPYKVSXGKIPTEGLMVDFS DIQPYGNHMGNSAPSVEADN					
	orf15-1	IKPKTNAFEAAAYKENYALWMGPYKVSXGKIPTEGLMVDFS DIRPYGNHTGNSAPSVEADN					
		250	260	270	280	290	300
10	orf15a.pep	SHEGYGYSDEAVRRHRQGPX					
	orf15-1	SHEGYGYSDEVVRQHRQGPX					
		310	320				

Further work identified the corresponding gene in *N.gonorrhoeae* <SEQ ID 83>:

15	1	ATGCGGGCAC	GGCTGCTGAT	ACCTATTCTT	TTTTCAGTTT	TTATTTTATC
	51	CGCCTGCGGG	ACACTGACAG	GTATTCCATC	GCATGGCGGA	GGCAAACGCT
	101	TCGCGGTCTGA	ACAAGAACTT	GTGGCCGCTT	CTGCCAGAGC	TGCCGTTAAA
	151	GACATGGATT	TACAGGCATT	ACACGGACGA	AAAGTTGCAT	TGTACATTGC
	201	AACATGGGC	GACCAAGGTT	CAGGCAGTTT	GACAGGGGGT	CGCTACTCCA
20	251	TTGATGCACT	GATTCGCGGC	GAATACATAA	ACAGCCCTGC	CGTCCGCACC
	301	GATTACACCT	ATCCGCGTTA	CGAAACCACC	GCTGAAACAA	CATCAGGCGG
	351	TTTGACGGGT	TTAACCCTT	CTTTATCTAC	ACTTAATGCC	CCTGCACCTCT
	401	CGCGCACCCA	ATCAGACGGT	AGCGGAAGTA	GGAGCAGTCT	GGGCTTAAAT
	451	ATTGGCGGGA	TGGGGGATTA	TCGAAATGAA	ACCTTGACGA	CCAACCCGCG
25	501	CGACACTGCC	TTTCTTTCCC	ACTTGGTGCA	GACCGTATTT	TTCCTGCGCG
	551	GCATAGACGT	TGTTTCTCCT	GCCAATGCCG	ATACAGATGT	GTTTATTAAAC
	601	ATCGACGTAT	TCGGAACGAT	ACGCAACAGA	ACCGAAATGC	ACCTATACAA
	651	TGCCGAAACA	CTGAAAGCCC	AAACAAAAC	GGAATATTTC	GCAGTAGACA
	701	GAACCAATAA	AAAATTGCTC	ATCAAACCCA	AAACCAATGC	GTTTGAAGCT
30	751	GCCTATAAAG	AAAATTACGC	ATTGTGGATG	GGGCCGTATA	AAGTAAGCAA
	801	AGGAATCAAA	CCGACGGAAG	GATTGATGGT	CGATTTCTCC	GATATCCAAC
	851	CATACGCAA	TCATACGGGT	AACTCCGCCC	CATCCGTAGA	GGCTGATAAC
	901	AGTCATGAGG	GGTATGGATA	CAGCGATGAA	GCAGTGCAC	AACATAGACA
	951	AGGGCAACCT	TGA			

35 This encodes a protein having amino acid sequence <SEQ ID 84; ORF15ng>:

	1	MRARLLIPIL	FSVFILSACG	TLTGIPSHGG	GKRFQVEQEL	VAASARAAVK
	51	DMDLQALHGR	KVALYIATMG	DQSGSLTGG	RYSIDALIRG	EYINSPAVRT
	101	DYTYPRYETT	AETTSGLTGT	LTTSLSTLNA	PALSRTQSDG	SGSRSSLGLN
	151	IGMGDYNRNE	TLTNPDRDTA	FLSHLVQTVF	FLRGIDVVSP	ANADTDVFIN
40	201	IDVFGTIRNR	TEMHLYNAET	LKAQTKLEYF	AVDRTNKKLL	IKPKTNAFEA
	251	AYKENYALWM	GPYKVSXGKIP	TEGLMVDFS	DIQPYGNHTG	NSAPSVEADN
	301	SHEGYGYSDE	AVRQHRQGP	*		

The originally-identified partial strain B sequence (ORF15) shows 97.2% identity over a 213aa overlap with ORF15ng:

45	orf15.pep	MQARLLIPILFSVFILSACGTLTGIPSHGGXKRFAVEQELVAASARAAVKDMDLQALHGR	60
	orf15ng	MRARLLIPILFSVFILSACGTLTGIPSHGGGKRFAVEQELVAASARAAVKDMDLQALHGR	60
	orf15.pep	KVALYIATMGDQSGSLTGGRYSIDAXXGEYINSPAVRTDYTYPRYETTAETTSGLTGT	120
50	orf15ng	KVALYIATMGDQSGSLTGGRYSIDALIRGEYINSPAVRTDYTYPRYETTAETTSGLTGT	120
	orf15.pep	LTTSLSTLNAPALSRTQSDGSGSKSSLNIGMGDYRNETLTTNPRDTAFLSHLVQTVF	180
55	orf15ng	LTTSLSTLNAPALSRTQSDGSGSRSSLNIGMGDYRNETLTTNPRDTAFLSHLVQTVF	180
	orf15.pep	FLRGIDVVSPANADTDVFINIDVFGTIRNRTEM	213
	orf15ng	FLRGIDVVSPANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL	240

60 The complete strain B sequence (ORF15-1) and ORF15ng show 98.8% identity in 320 aa overlap:

		10	20	30	40	50	60
	orf15-1.pep	MQARLLIPILFSVFILSACGTLTGIPSHGGGKRFAVEQELVAASARAANKDMDLQALHGR					
	orf15ng	MRARLLIPILFSVFILSACGTLTGIPSHGGGKRFAVEQELVAASARAANKDMDLQALHGR					
5		10	20	30	40	50	60
	orf15-1.pep	KVALYIATMGDQGSGLTGGRYSIDALIRGEYINSPAVRTDYTPRYETTAETSSGGLTG					
10	orf15ng	KVALYIATMGDQGSGLTGGRYSIDALIRGEYINSPAVRTDYTPRYETTAETSSGGLTG					
		70	80	90	100	110	120
	orf15-1.pep	LTTSLSTLNAPALSRTQSDGSGSKSLGLNIGMGDYRNETLTNNPRDTAFLSHLVQTVF					
15	orf15ng	LTTSLSTLNAPALSRTQSDGSGSRSLGLNIGMGDYRNETLTNNPRDTAFLSHLVQTVF					
		130	140	150	160	170	180
	orf15-1.pep	FLRGIDVVSANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL					
20	orf15ng	FLRGIDVVSANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL					
		190	200	210	220	230	240
	orf15-1.pep	IKPKTNAFEAAAYKENYALWMGPYKVSIGIKPTEGLMVDSDIRPYGNHTGNSAPSVEADN					
25	orf15ng	IKPKTNAFEAAAYKENYALWMGPYKVSIGIKPTEGLMVDSDIQPYGNHTGNSAPSVEADN					
		250	260	270	280	290	300
	orf15-1.pep	SHEGYGYSDEVVRQHRQGQPX					
30	orf15ng	SHEGYGYSDEAVRQHRQGQPX					
		310	320				
35							

Computer analysis of these amino acid sequences reveals an ILSAC motif (putative membrane lipoprotein lipid attachment site, as predicted by the MOTIFS program).

indicates a putative leader sequence, and it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF15-1 (31.7kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 4A shows the results of affinity purification of the GST-fusion protein, and Figure 4B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 4C) and ELISA (positive result). These experiments confirm that ORFX-1 is a surface-exposed protein, and that it is a useful immunogen.

#### Example 11

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 85>:

1 ..GG.CAGCACA AAAACAGGC GGTGAACGG AAAAACCCTA TTTACGATGA  
51 TGCCGGGTAT GATATTCGGC GTATTCACGG GCGCATTCTC CGCAAAATAT

5  
10  
101 ATCCCCGCGT TCGGGCTTCA AATTTTCTTC ATCCTGTTTT TAACCGCCGT  
151 CGCATTCAAA AACTGTCATA CCGACCCTCA GACGGCATCC CGCCCGCTGC  
201 CCGGACTGCC C<sub>2</sub>ACTGACT GCGGTTTCCA CACTGTTCGG CACAATGTGC  
251 AGCTGGGTCG GCATAGGCGG CGGTTCACTT TCCGTCCCCT TCTTAATCCA  
301 CTGCGGCTTC CCCGCCATA AAGCCATCGG CACATCATCC GGCCTTGCCCT  
351 GGCCGATTGC ACTCTCCGGC GCAATATCGT ATCTGCTCAA CGGCCTGAAT  
401 ATTGCAAGGAT TGCCCGAAGG GTCCTGGGC TTCCTTTACC TGCCCGCCGT  
451 CGCCGTCCTC AGCGCGGCAA CCATGTCCTT TGCCCGGCTC GGTGTCAAAA  
501 CCGCCACAA ACTTTCTTCT GCCAACTCA AAAAATC.TT CGGCATTATG  
551 TTGCTTTTGA TTGCCGGAAT AATGCTGTAC AACCTGCTTT AA

This corresponds to the amino acid sequence <SEQ ID 86; ORF17>:

1 ..GQHKQAVNG KTVFTMPPGM IFGVFTGAFS AKYIPAFGLQ IFFILELTAV  
51 AFKTLHTDPQ TASRPLPGLP XLTAVSTLFG TMSSWVGIGG GSLSVPFLIH  
101 CGFPAHKAIG TSSGLAWPIA LSGAISYLLN GLNIAGLP<sub>EG</sub> SLGFLYLP<sub>AV</sub>  
151 AVL<sub>SA</sub>ATIAF APLGVKTAHK LSSAKLK<sub>KS</sub>F GIMLLLIAGK MLYNLL\*

Further work revealed the complete nucleotide sequence <SEQ ID 87>:

20  
25  
30  
35  
1 ATGTGGCATT GGGACATTAT CTTAATCCTG CTTGCCGTAG GCAGTGCGGC  
51 AGGTTTTTATT GCCGGCCTGT TCGGCGTAGG CGGCGGCACG CTGATTGTCC  
101 CTGTGCTTTT ATGGGTGCTT GATTTCAGG GTTTGGCACA ACATCCTTAC  
151 GCGCAACACC TCGCCGTCGG CACATCCTTC GCCGTCATGG TCTTCACCGC  
201 CTTTCCAGT ATGCTGGGGC AGCACAAAAA ACAGGCGGTC GACTGGAAAA  
251 CCGTATTTAC GATGATGCCG GGTATGATAT TCGGCGTATT CACGGGCGCA  
301 CTCTCCGCAA AATATATCCC CGCGTTCGGG CTTCAAATTT TCTTCATCCT  
351 GTTTTAAACC GCCGTCGCAT TCAAAACACT GCATACCGAC CCTCAGACGG  
401 CATCCCGCCC GCTGCCCGGA CTGCCCGGAC TGA<sub>CT</sub>GCGGT TTCCACACTG  
451 TTCGGCACAA TGTCGAGCTG GGTCCGCATA GCGGCGGTT CACTTCCGT  
501 CCCCTTCTTA ATCCACTGCG GCTTCCCGC CCATAAGCC ATCGGCACAT  
551 CATCCGGCCT TGCTGGCCG ATTGCACTCT CCGGCGCAAT ATCGTATCTG  
601 CTCAACGGCC TGAATATGCG AGGATTGCCC GAAGGGTCAC TGGGCTTCCT  
651 TTACCTGCCC GCGGTCGCGC TCCTCAGCGC GGCAACCAT TGCCTTGCCC  
701 CGCTCGGTGT CAAAACCGCC CACAACTTT CTTCTGCCAA ACTCAAAAA  
751 Tc.TTCGGCA TTATGTTGCT TTTGATTGCC GGAAAAATGC TGTACAACT  
801 GCTTTAA

This corresponds to the amino acid sequence <SEQ ID 88; ORF17-1>:

35  
40  
1 MWHWDIILIL LAVGSAAGFI AGLEFGVGGT LIVPVVLWVL DLOGLAQHPY  
51 AQHLAVGTSE AVMVETAFSS MLGQHKQAV DWKTVFTMMP GMIFGVETGA  
101 LSAKYIPAFG LQIFFILELT AVAFKTLHTD PQTASRPLPG LPGLTAVSTL  
151 FGTMSSWVGI GGSLSVPFL IHCGFPAHKA IGTSSGLAWP IALSGAISYL  
201 LNGLNIAGLP EGSGLFLYLP AVAVLSAATI AFAPLVKTA HKLSSAKLKK  
251 XFGIMLLLIA GKMLYNLL\*

Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical *H.influenzae* transmembrane protein HI0902 (accession number P44070)

ORF17 and HI0902 proteins show 28% aa identity in 192 aa overlap:

45  
50  
55  
ORF17 3 HKKQAVNGKTVFTMPPGMIFGVET-GAFSAKYIPAFGLQIF--FILELTAVAFKTLHTDP 59  
HK + + V + P ++ VF G F + +IF +++L ++ D  
HI0902 72 HKLGNIVWQAVRILAPVIMLSVFICGLFIGRLDREISAKIFACLVVYLATKMVLSIKKD- 130  
ORF17 60 QTASRPLPGLPXLTAVSTLFGTMSSWVGIGGGSLSVPFLIHCGFPAHKAIGTSSGLAWPI 119  
Q ++ L L + L G SS GIGGG VPFL G +AIG+S+ +  
HI0902 131 QVTTKSLTPLSSVIG-GILIGMASSAGIGGGFIVPFLTARGINIKQAIGSSAFCGMLL 189  
ORF17 120 ALSGAISYLLNGLNIAGLPEGSLGFLYLP<sub>AV</sub>AVLSAATIAFAPLVGXXXXXXXXXXXXX 179  
+SG S+++G +PE SLG++YLP<sub>AV</sub> ++A + + LG  
HI0902 190 GISGMFSFIVSGWGNPLMPEYSLGYIYLP<sub>AV</sub>LGITATSFFTSKLGASATAKLPVSTLKKG 249  
ORF17 180 FGIMLLLIAGKM 191  
F + L+++A M  
HI0902 250 FALFLIVVAINM 261



Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF17 shows 96.9% identity over a 196aa overlap with an ORF (ORF17a) from strain A of *N. meningitidis*:

```

5      orf17.pep      10      20      30
                        GQHKHKQAVNGKTVFTMMPGMIFGVFTGAFS
                        |||||:|||||:|||||:|
      orf17a      QGLAQHPYAQHLAVGTSFAVMVFTAFSSMLGQHKHKQAVDWKTVFTMMPGMVFGVFAGALS
                        50      60      70      80      90      100

10     40      50      60      70      80      90
      orf17.pep      AKYIPAFGLQIIFFILELTAVAFKTLHTDPQTASRPLPGLPXLTAVSTLFGTMSSSWVGIGG
                        |||||:|||||:|||||:|
      orf17a      AKYIPAFGLQIIFFILELTAVAFKTLHTDPQTASRPLPGLPGLTAVSTLFGTMSSSWVGIGG
                        110     120     130     140     150     160

15     100     110     120     130     140     150
      orf17.pep      GSLSVPFLIHCGFPAHKAIGTSSGLAWPIALSGAISYLLNGLNIAGLPEGSLGFLYLPVAV
                        |||||:|||||:|||||:|
      orf17a      GSLSVPFLIHCGFPAHKAIGTSSGLAWPIALSGAISYLLNGLNIAGLPEGSLGFLYLPVAV
                        170     180     190     200     210     220

20     160     170     180     190
      orf17.pep      AVLSAATIAFAPLGVKTAHKLSSAKLKKSFGIMLLLIAGKMLYNLLX
                        |||||:|||||:|||||:|
      orf17a      AVLSAATIAFAPLGVKTAHKLSSAKLKKSFGIMLLLIAGKMLYNLLX
                        230     240     250     260

```

The complete length ORF17a nucleotide sequence <SEQ ID 89> is:

	1	ATGTGGCATT	GGGACATTAT	CTTAATCCTG	CTTGCCGTAG	GCAGTGCGCC
30	51	AGGTTTATT	GCCGGCCTGT	TCGGCGTAGG	CGGCGGCACG	CTGATTGTCC
	101	CTGTCGTTTT	ATGGGTGCTT	GATTTCAGG	GTTTGGCACA	ACATCCTTAC
	151	CGCGAACACC	TCGGCGTCGG	CACATCCTTC	GCCGCTCATTG	TCTTCACCGC
	201	CTTTTCCAGT	ATGCTGGGGC	AGCACAAAAA	ACAGGCGGTC	GACTGGAAAA
	251	CCGTATTTAC	GATGATGCCG	GGTATGGTAT	TCCGCGGTAT	CGCTGGCGCA
35	301	CTCTCCGCAA	AATATATCCC	AGCGTTCGGG	CTTCAAATTT	TCTTCATCCT
	351	GTTTTTAACC	GCCGTCGCAT	TCAAAACACT	GCATACCGAC	CCTCAGACGG
	401	CATCCCGCCC	GCTGCCCGGA	CTGCCCGGAC	TGACTTCGGT	TCTCCACTCG
	451	TTCGGCACAA	TGTCGAGCTG	GGTCGGCATA	GGCGCGGGTT	TTATTTCCGT
	501	CCCCTTCTTA	ATCCACTGCG	GCTTCCCCGC	CCATAAAGCC	ATCGGCACAT
40	551	CATCCGGCCT	TGCCTGGCCG	ATTGCACTCT	CCGGCGCAAT	ATCGTATCTG
	601	CTCAACGGCC	TGAATATTGC	AGGATTGCCC	GAGGGGTCAC	TGGGCTTCTT
	651	TTACTTCCCC	GCCGTCGCCG	TCTCAGCGC	GGCAACCAT	GCCTTTGCCC
	701	CGCTCGGTGT	CAAAACGCC	CACAAACTTT	CTTCTGCCAA	ACTCAAAAAA
	751	TCCTTCGGCA	TTATGTTGCT	TTTGATTGCC	GGAAAAATGC	TGTACAACCT
	801	GCTTTAA				

45 This encodes a protein having amino acid sequence <SEQ ID 90>:

50

1	MWHWDIILIL	LAVGSAAGFI	AGLFGVGGGT	LIVPVVLWVL	DLOGLAQHPY
51	AQHLAGVTSF	AVMVFTAFSS	MLGQHKKQAV	DWKTVFTMMP	MGVFGVFAGA
101	LSAKYIPAFG	IQIFFILFLT	AVAETKLHTD	PQTASRPLPG	LPGLTAVSTL
151	FGTMSSWVGI	GGGSLVFPFL	IHCGFPAHKA	IGTSSGLAWP	IALSGAISYL
201	LNGLNIAGLP	EGSLGFLYLP	AVAVLSAATI	AFAPLGVKTA	HKLSSAKLKK
251	SFGIMLLLIA	GKMLYNLL*			

ORF17a and ORF17-1 show 98.9% identity in 268 aa overlap:

		10	20	30	40	50	60
<b>55</b>	orf17a.pep	MWHWDIILILLAVGSAAGFIAGLFGVGGGTTLIVPVLWVLDLQGLAQHPYAQHHLAVGTSTF 					
	orf17-1	MWHWDIILILLAVGSAAGFIAGLFGVGGGTTLIVPVLWVLDLQGLAQHPYAQHHLAVGTSTF 					
		10	20	30	40	50	60
<b>60</b>	orf17a.pep	AVMVFATAFSMMLGQHKQQAVDWKTVFTMMPGMVFGVFAALSAKYIPAFGLQIFFILFLT 					
			70	80	90	100	110

		:   :   :
	orf17-1	AVMVFTAFSSMLGQHKKQAVDWKTVFTMMPGMIFGVFTGALSAKYIPAFGLQIFFILFLT
		70 80 90 100 110 120
5		130 140 150 160 170 180
	orf17a.pep	AVAFKTLHTDPQTASRPLPGLPGLTAVSTLFGTMSSWVGIGGGSLSVPFLLHCGFPAHKA
	orf17-1	AVAFKTLHTDPQTASRPLPGLPGLTAVSTLFGTMSSWVGIGGGSLSVPFLLHCGFPAHKA
		130 140 150 160 170 180
10		190 200 210 220 230 240
	orf17a.pep	IGTSSGLAWPIALSGAISYLLNGLNIAGLPEGSLGFLYLPAAVLASAATIAFAPLGVKTA
	orf17-1	IGTSSGLAWPIALSGAISYLLNGLNIAGLPEGSLGFLYLPAAVLASAATIAFAPLGVKTA
15		190 200 210 220 230 240
		250 260 269
	orf17a.pep	HKLSSAKLKKSFGIMLLLIAGKMLYNLLX
20	orf17-1	HKLSSAKLKXFGIMLLLIAGKMLYNLLX
		250 260

Homology with a predicted ORF from *N.gonorrhoeae*

ORF17 shows 93.9% identity over a 196aa overlap with a predicted ORF (ORF17.ng) from *N.*

*gonorrhoeae:*

	orf17.pep	GQHKKQAVNGKTVFTMMPGMIFGVFTGAFS	30
	orf17.ng	:   : :       : :	
	orf17.ng	QGLAQHPYAQHLAVGTSFAVMVTFATSSMLGQHKKQAVDWKTI FAMPMPGMIFGVFAGALS	102
30	orf17.pep	AKYIPAFGLQIFFILFLTAVAFKTLTHTDPQTASRPLPGLPXLTAVSTLFGTMSWVGIGG	90
	orf17.ng	:   : :       : :	
	orf17.ng	AKYIPAFGLQIFFILFLTAVAFKTLTHTGRQTASRPLPGLPGLTAVSTLF GAMSWSWVGIGG	162
	orf17.pep	GSLSVPFLIHCGFPAHKAIGTSSGLAWPIALSGAISYLLNGLNIAGLPEGSLGFLYLP	150
35	orf17.ng	:   : :       : :	
	orf17.ng	GSLSVPFLIHCGFPAHKAIGTSSGLAWPIALSGAISYLVNGLNIAGLPEGSLGFLYLP	202
	orf17.pep	AVLSAATIAFAPLGVKTAHKLSSAKLKKSFGIMLLLIAGKMLYNLL	196
40	orf17.ng	:   : :       : :	
	orf17.ng	AVLSAATIAFAPLGVKTAHKLSSAKLKESFGIMLLLIAGKMLYNLL	268

An ORF17ng nucleotide sequence <SEQ ID 91> is predicted to encode a protein having amino acid sequence <SEQ ID 92>:

45

1	MWHWDIILIL	LAVGSAAGFI	AGLFGVGGGT	LIVPVVLWVL	DLQGLAQHPY
51	AQHLAVGTFSF	AVMVFTAFSS	MLGQHKKQAV	DWKTIFAMFP	GMIFGVFAGA
101	LSAKYIPAFG	QIQQIFILELT	AVAFKTLHTG	RQTASRLPG	LPGLTAVSTL
151	FGAMSSWVGI	GGGSLSVPFL	IHCGFPAHKA	IGTSSGLAWP	IALSGAISYL
201	VNGLNIAGLP	EGSLGLFYLP	AVAVLSAATI	AFAPLGVKTA	HKLSSAKLKE
251	SFGIMLLLIA	GKMLYNLL*			

Further work revealed the complete gonococcal DNA sequence <SEQ ID 93>:

50	1	ATGTGGCATT	GGGACATTAT	CTTAATCTCG	CTTGCgtag	gcAGTGCGGC
	51	AGGTTTTATT	GCCGGCCTGT	Tcgggtgtagg	cggcgGTACG	CTGATTTGTCC
	101	CTGTCTGTTTT	ATGGGTCGTT	GATTTGCAGG	gTTTGGGACA	ACATCCTTAC
	151	GCGCAACACC	TGCGCGTCGG	CAcaTccttc	gcCGTCATGG	TCTTCCACGG
55	201	CTTTTCCAGT	ATGTTGGGGC	AGCACAAAAA	ACAGGCGGTC	GACTGGA AAA
	251	CCATATTTGC	GATGATCCCG	GGTATGATAT	TCGGCGTATT	CGCTGGCGCA
	301	CTCTCCGCAA	AATATATCCC	CGCGTTCGGG	CTTCAAAATT	TCTTCACTCT
	351	GTTTTTAACC	GCGCTCGCAT	TCAAAACACT	GCATACCGGT	CGTCAGACGG
60	401	CATCCGCCCC	GCTGCCCGGG	CTGCCCGGAC	TGACTCGGGT	TTCCACACTG
	451	TTCCGCGCAA	TGTCGAGCTG	GGTCGGGCAT	GGCGGCGGTT	CAC TTTCCGT
	501	CCCCTTCTTA	ATCCACTGCG	GCTTCCCGCG	CCATAAAGCC	ATCGGCACAT

5 551 CATCCGGCCT TGCCTGGCCG ATTGCACTCT CCGGCGCAAT ATCGTATCTG  
 601 GTCAACGGTC TGAATATTGC AGGATTGCCC GAAGGTCGCG TGGGCTTCCT  
 651 TTACCTGCCC GCCGTCGCCG TCCTCAGCGC GGCAACCATT GCCTTTGCC  
 701 CGCTCGGTGT CAAAACCGCC CACAACTTT CTTCTGCCAA ACTCAAAGAA  
 751 TCCTTCGGCA TTATGTTGCT TTTGATTGCC GGAAAAATGC TGTACAACCT  
 801 GCTTTAA

This corresponds to the amino acid sequence <SEQ ID 94; ORF17ng-1>:

10 1 MWHWDIILIL LAVGSAAGFI AGLFGVGGT LIVPVVLWVL DLQGLAQHPY  
 51 AQHLAVGTSF AVMVFTAFSS MLGQHKQAV DWKTIFAMMP GMIFGVFAGA  
 101 LSAKYIPAFG LQIFFILFLT AVAFKTLHTG RQTASRPLPG LPGLTAVSTL  
 151 FGAMSSWVG I GGGSLVPFL IHCGFPAHKA IGTSSGLAWP IALSGAISYL  
 201 VNGLNIAGLP EGSLGFLYLP AVAVLSAATI AFAPLVKTA HKLSSAKLKE  
 251 SFGIMLLLIA GKMLYNLL\*

ORF17ng-1 and ORF17-1 show 96.6% identity in 268 aa overlap:

15 orf17-1.pep 10 20 30 40 50 60  
 MWHWDIILILLAVGSAAGFIAGLFGVGGTLIVPVVLWVLDLQGLAQHPYAQHLAVGTSF  
 orf17ng-1 MWHWDIILILLAVGSAAGFIAGLFGVGGTLIVPVVLWVLDLQGLAQHPYAQHLAVGTSF  
 20 10 20 30 40 50 60  
 orf17-1.pep 70 80 90 100 110 120  
 AVMVFTAFSSMLGQHKQAVDWKTIVFTMPGMIFGVFTGALS AKYIPAFGLQIFFILFLT  
 orf17ng-1 AVMVFTAFSSMLGQHKQAVDWKTIFAMMPGMIFGVFAGALS AKYIPAFGLQIFFILFLT  
 25 70 80 90 100 110 120  
 orf17-1.pep 130 140 150 160 170 180  
 AVAFKTLHTDPQTASRPLPGLTAVSTLFGTMSSWVGIGGSLVPFLIHCGFPAHKA  
 orf17ng-1 AVAFKTLHTGRQTASRPLPGLTAVSTLFGAMSSWVGIGGSLVPFLIHCGFPAHKA  
 30 130 140 150 160 170 180  
 orf17-1.pep 190 200 210 220 230 240  
 IGTSSGLAWPIALSGAISYLLNGLNIAGLPEGSLGFLYLP AVAVLSAATIAFAPLVKTA  
 orf17ng-1 IGTSSGLAWPIALSGAISYLVNGLNIAGLPEGSLGFLYLP AVAVLSAATIAFAPLVKTA  
 35 190 200 210 220 230 240  
 orf17-1.pep 250 260 269  
 HKLSSAKLKKXFGIMLLIAGKMLYNLLX  
 orf17ng-1 HKLSSAKLKESFGIMLLIAGKMLYNLLX  
 40 250 260

In addition, ORF17ng-1 shows significant homology with a hypothetical *H. influenzae* protein:

45 sp|P44070|Y902\_HAEIN HYPOTHETICAL PROTEIN HI0902 pir||G64015 hypothetical protein  
 HI0902 - Haemophilus influenzae (strain Rd KW20) gi|1573922 (U32772) H. influenzae  
 predicted coding region HI0902 [Haemophilus influenzae]Length = 264  
 Score = 74 (34.9 bits), Expect = 1.6e-23, Sum P(2) = 1.6e-23  
 Identities = 15/43 (34%), Positives = 23/43 (53%)  
 50 Query: 55 AVGTSFAVMVFTAFSSMLGQHKQAVDWKTIFAMMPGMIFGVF 97  
 A+GTSFA +V T S HK + W+ + + P ++ VF  
 Sbjet: 52 ALGTSFATIVITGIGSAQRHHKLGNIWQAVRILAPVIMLSVF 94  
 55 Score = 195 (91.9 bits), Expect = 1.6e-23, Sum P(2) = 1.6e-23  
 Identities = 44/114 (38%), Positives = 65/114 (57%)  
 Query: 150 LFGAMSSWVGIGGSLVPFLIHCGFPAHKAIGTSSGLAWPIALSGAISYLVNGLNIAGL 209  
 L G SS GIGGG VPFL G +AIG+S+ + +SG S++V+G +  
 60 Sbjet: 148 LIGMASSAAGIGGGFIVPFLTARGINIKQAIGSSAFCGMLLGISGMFSFIVSGWGNPLM 207  
 Query: 210 PEGSLGFLYLP AVAVLSAATIAFAPLVKTA HKLSSAKLKESFGIMLLIAGKM 263  
 PE SLG++YLP AV ++A + + LG KL + LK+ F + L+++A M

Sbjct: 208 PEYSLGYIYLPVAVLGITATSFFTSKLGASATAKLPVSTLKKGFALFLIVVAINM 261

This analysis, including the homology with the hypothetical *H.influenzae* transmembrane protein, suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 12

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 95>:

```

10      1  ..GGAAACGGAT GGCAGGCAGA CCCCGAACAT CCGCTGCTCG GGCTTTTTCG
      51  CGTCAGTAAT GTATCGATGA CGCTTGCTTT TGTCGGAATA TGTGCGTTGG
     101  TGCATTATTG CTTTTCGGGA ACGGTTCAAG TGTTTGTGTT TGCGGCACTG
     151  CTCAAACTTT ATGCGCTGAA GCCGGTTTAT TGGTTCGTGT TGCAGTTTGT
     201  GCTGATGGCG GTTGCCTATG TCCACCGCTG CGGTATAGAC CGGCAGCCGC
     251  CGTCAACGTT CGGCGGCTCG CAGCTGCGAC TCGGCGGGTT GACGCAGCG
     301  TTGATGCAGG TCTCGGTACT GGTGCTGCTG CTTTCAGAAA TTGGAAGATA
     351  A
  
```

This corresponds to the amino acid sequence <SEQ ID 96; ORF18>:

```

20      1  ..GNGWQADPEH PLLGLFAVSN VSMTLAFVGI CALVHYCFSG TVQVFVFAAL
     51  LKLYALKPVY WFLVQLFVMA VAYVHRCGID RQPPSTFGGS QLRLLGLTAA
    101  LMQVSVLVLL LSEIGR*
  
```

Further work revealed the complete nucleotide sequence <SEQ ID 97>:

```

25      1  ATGATTTTGC TGCATTGGA TTTTTGTCT GCCTTACTGT ATGCGGCGGT
     51  TTTTCTGTTT CTGATATTCC GCGCAGGAAT GTTGCAATGG TTTTGGGCGA
    101  GTATTATGCT GTGGCTGGGC ATATCGGTTT TGGGGGCAAA GCTGATGCCC
    151  GGCATATGGG GAATGACCCG CGCCGCGCCC TTGTTTCATCC CCCATTTTTA
    201  CCTGACTTTG GGCAGCATAT TTTTTTTCAT CGGGCATTGG AACCGGAAAA
    251  CAGATGGAAA CGGATGGCAG GCAGACCCCG AACATCCGCT GCTCGGGCTT
    301  TTTGCCGTCA GTAATGTATC GATGACGCTT GCTTTTGTCT GAATATGTGC
    351  GTTGGTG CAT TATTGCTTTT CGGGAACGGT TCAAGTGTTT GTGTTTGGCG
    401  CACTGCTCAA ACTTTATGCG CTGAAGCCGG TTTATTGTTT CGTGTTGCAG
    451  TTTGTGCTGA TGGCGGTTGC CTATGTCCAC CGCTGCGGTA TAGACCGGCA
    501  GCCGCCGTCA ACGTTCGGCG GCTCGCAGCT GCGACTCGGC GGGTTGACGG
    551  CAGCGTTGAT GCAGGTCTCG GTACTGGTGC TGCTGCTTTC AGAAATTGGA
    601  AGATAA
  
```

This corresponds to the amino acid sequence <SEQ ID 98; ORF18-1>:

```

40      1  MILLHLDFLS ALLYAAVFLF LIFRAGMLQW FWASIMLWLG ISVLGAKLMP
     51  GIWGMTRAAP LFIPHFYLTLSGIFFFIHWN NRKTDGNGWQ ADPEHPLLGL
    101  FAVSNVSMFL AFGICALVH YCFSGTVQVF VFAALLKLYA LKPVYWFVLQ
    151  FVLMAVAVYH RCGIDRQPPS TFGGSQRLRG GLTAALMQVS VLVLLLSEIG
    201  R*
  
```

Computer analysis of this amino acid sequence gave the following results:

### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF18 shows 98.3% identity over a 116aa overlap with an ORF (ORF18a) from strain A of *N.meningitidis*:

```

45      orf18.pep                                10      20      30
      orf18a      TRAAPLFIPHFYLTLSGIFFFIHWNRKTDGNGWQADPEHPLLGLFAVSNVSMTLAFVGI
                  60      70      80      90      100     110
  
```

		40	50	60	70	80	90
	orf18.pep	<u>CALVHYCFSGTVQVFVFAALLKLYALKPVYWFVLQFVLMAYVHRCGIDRQPPSTFGGS</u>					
5	orf18a						
		<u>CALVHYCFSGTVQVFVFAALLKLYALKPVYWFVLQFVLMAYVHRCGIDRQPPSTFGGS</u>					
		120	130	140	150	160	170
		100	110				
10	orf18.pep	<u>QLRLGGLTAALMQVSVLVLLSEIGRX</u>					
	orf18a						
		<u>QLRLGGLTAALMQVSVLVLLSEIGRX</u>					
		180	190	200			

The complete length ORF18a nucleotide sequence <SEQ ID 99> is:

1	ATGATTTTGC	TGCATTTGGA	TTTTTTGTCT	GCCTTACTGT	ATGCGGCGGT
15	51	TTTTCTGTTT	CTGATATGCC	GCGCAGGAAT	GTTGCAATGG
	101	GTATTATGCT	GTGGCTGGGC	ATATCGGTTT	TGGGGGCAAA
	151	GGCATATGGG	GAATGACCCG	CGCCGCGCCC	TTGTTTCATCC
	201	CCTGACTTTG	GGCAGCATAT	TTTTTTTCAT	CGGGCATTGG
	251	CGGATGGAAA	CGGATGGCAG	GCAGACCCCG	AACATCCTCT
20	301	TTTGCCGTCA	GTAATGTATC	GATGACGCTT	GCTTTTGTCT
	351	GTTGGTGCAT	TATTGCTTTT	CGNGAACGGT	TCAAGTGTTC
	401	CACTGCTCAA	ACTTTATGCG	CTGAAGCCGG	TTTATTGGTT
	451	TTTGTGCTGA	TGGCGGTTGC	CTATGTCCAC	CGCTGCGGTA
	501	GCCGCGGTCA	ACGTTGCGCG	GNTCGCAGCT	GCGACTCGGC
25	551	CAGCGTTGAT	GCAGNTCTCG	GTACTGGTGC	TGCTGCTTTC
	601	AGATAA			

This encodes a protein having amino acid sequence <SEQ ID 100>:

1	MILLHLDFLS	ALLYAAVFLF	LIFRAGMLQW	FWASIMLWL	ISVLGAKLMP
51	GIWGMTRAAP	LFIPHFYLT	LSIFFFIHGW	NRKTDGNGWQ	ADPEHPLLGL
101	FAVSNVSMTL	AFVGICALVH	YCFSTVQVF	VFAALLKLYA	LKPVYWFVLQ
151	FVLMAYAVVH	RCGIDRQPPS	TFGGSQLRLG	GLTAALMQXS	VLVLLSEIG
201	R*				

ORF18a and ORF18-1 show 99.0% identity in 201 aa overlap:

35	orf18a.pep	10	20	30	40	50	60
		MILLHLDFLSALLYAAVFLFLIFRAGMLQWF	WASIMLWLGISVLGAKLMPGIWGMTRAAP				
	orf18-1	MILLHLDFLSALLYAAVFLFLIFRAGMLQWF	WASIMLWLGISVLGAKLMPGIWGMTRAAP				
40	orf18a.pep	70	80	90	100	110	120
		LFIPHFYLTLGSIFFFI	GHWNRKTDGNGWQADPEHPLLGLFAVSNVSM	TLAFVGICALVH			
	orf18-1	LFIPHFYLTLGSIFFFI	GHWNRKTDGNGWQADPEHPLLGLFAVSNVSM	TLAFVGICALVH			
45	orf18a.pep	130	140	150	160	170	180
		YCF	SXTVQVF	FAALLKLYALKPVYWFVLQFVLM	AVAYVHRCGIDRQPPST	FFGGSQQLRLG	
	orf18-1	YCF	SGTVQVF	FAALLKLYALKPVYWFVLQFVLM	AVAYVHRCGIDRQPPST	FFGGSQQLRLG	
50	orf18a.pep	190	200				
		GLTAALMQXSVLVLL	SEIGRX				
	orf18-1	GLTAALMQXSVLVLL	SEIGRX				
55	orf18a.pep	190	200				
		GLTAALMQXSVLVLL	SEIGRX				
	orf18-1	GLTAALMQXSVLVLL	SEIGRX				

Homology with a predicted ORF from *N.gonorrhoeae*

ORF18 shows 93.1% identity over a 116aa overlap with a predicted ORF (ORF18.ng) from *N. gonorrhoeae*:

	orf18.pep		GNGWQADPEHPLLGLFAVSNVSMTLAFVGI	30
	orf18ng	TRAAPLFIPHFYLTLSIFFFIGYWNKTDGNGWQADPEHPLLGLFAVSNVSMTLAFVGI		115
5	orf18.pep	CALVHYCFSGTVQVFVFAALLKLYALKPVYWFVLQFVLMAYVHRCGIDRQPPSTFGGS		90
	orf18ng	CALVHYCFSGTVQVFVFAALLKLYALKPVYWFVLQFVLMAYVHRCGIDRQPPSTFGGS		175
	orf18.pep	QLRLGGLTAALMQVSVLVLLSEIGR	116	
10		::      :  ::  :		
	orf18ng	QLRLGVLAAMLQVAVTAMLLAEIGR	201	

The complete length ORF18ng nucleotide sequence is <SEQ ID 101>:

	1	ATGATTTTGC	TGCATTTGGA	TTTTTTGTCT	GCCTTACTGt	aTGCGGcggt
	51	tttTctgTTT	CTGATATTCC	GCGCAGGAAT	GTGCAATGG	TTTTGGGCGA
15	101	GTATTGCGTT	GTGGCTCGGC	ATCTCGGTTT	TAGGGGTAAA	GCTGATGCCG
	151	GGGATGTGGG	GAATGACCCG	CGCCGCGCCT	TTGTTTCATCC	CCCATTTTTA
	201	CCTGACTTTG	GGCAGCATAT	TTTTTTTCAT	CGGGTATTGG	AACCGGAAAA
	251	CAGATGAAAA	CGGATGGCAG	GCAGACCCCG	AACATCCGCT	GCTCGGGCTT
	301	TTTGCCGTCA	GTAATGTATC	GATGACGCTT	GCTTTTGTCT	GAATATGTGC
20	351	GTGGTGTCAT	TATTGCTTTT	CGGGAACGGT	TCAAGTGTTT	GTGTTTGC GG
	401	CATTGCTCAA	ACTTTATGCG	CTGAAGCCGG	TTTATTGGTT	CGTGTTCAG
	451	TTTGTATGA	TGGCGGttgC	CTATGTCCAC	CGCTGCGGTA	TAGACCGGCA
	501	GCCGCCGTCA	ACGTTCCGCG	GTTCCGAGCT	GCGACTCGGC	GTGTTGGCGG
	551	CGATGTTGAT	GCAGGTTGCG	GTAACGGCGA	TGCTGCTTGC	CGAAATCGGC
25	601	AGATGA				

This encodes a protein having amino acid sequence <SEQ ID 102>:

	1	MILLHLDFLS	ALLYAAVFLF	LIFRAGMLQW	FWASIALWLG	ISVLGVKLMP
	51	GMWGMTRAAP	LFIPHFYLT	LSIFFFIGYW	NRKTDGNGWQ	ADPEHPLLGL
30	101	FAVSNVSMTL	AFVGICALVH	YCFSGTVQVF	VFAALLKLYA	LKPVYWFVLQ
	151	FVLMAYAVH	RCGIDRQPPS	TFGGSQRLRG	VLAAMLQVA	VTAMLLAEIG
	201	R*				

This ORF18ng protein sequence shows 94.0% identity in 201 aa overlap with ORF18-1:

		10	20	30	40	50	60
35	orf18-1.pep	MILLHLDFLSALLYAAVFLFLIFRAGMLQWFWASIMLWLGISVLGAKIMPGIWMTRAAP					
	orf18ng	MILLHLDFLSALLYAAVFLFLIFRAGMLQWFWASIALWLGISVLGVKLMPGMWGMTRAAP					
		10	20	30	40	50	60
40	orf18-1.pep	LFIPHFYLTLSIFFFIGHWNKTDGNGWQADPEHPLLGLFAVSNVSMTLAFVGICALVH					
	orf18ng	LFIPHFYLTLSIFFFIGYWNKTDGNGWQADPEHPLLGLFAVSNVSMTLAFVGICALVH					
		70	80	90	100	110	120
45	orf18-1.pep	YCFSGTVQVFVFAALLKLYALKPVYWFVLQFVLMAYAVHRCGIDRQPPSTFGGSQRLRG					
	orf18ng	YCFSGTVQVFVFAALLKLYALKPVYWFVLQFVLMAYAVHRCGIDRQPPSTFGGSQRLRG					
		130	140	150	160	170	180
50	orf18-1.pep	GLTAALMQVSVLVLLSEIGRX					
	orf18ng	VLAAMLQVAVTAMLLAEIGRX					
55		190	200				

Based on this analysis, including the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

**Example 13**

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 103>:

```
1 ATGAAAACCC CACTCCTCAA GCCTCTGCTN ATTACCTCGC TTCCCCTTTT
5 51 CGCCAGTGTT TTTACCGCCG CCTCCATCGT CTGGCAGCTA GGCGAACCCA
101 AGCTCGCCAT GCCCTTCGTA CTCGGCATCA TCGCCGGCGG CTTTGTGGAT
151 TTGGACAACC NCNTGACCGG ACGGCTNAAA AACATCATCA CCACCGTCGC
201 CCTGTTCAACC CTCTCCTCGC TCACGGCACA AAGCACCTC GGCACAGGGC
251 TGCCCTTCAT CCTCGCCATG ACCCTGATGA CTT.CG.CTT CACCATTTTA
301 GGCGCGNCG ...
```

10 This corresponds to the amino acid sequence <SEQ ID 104; ORF19>:

```
1 MKTPLLKPLL ITSLPVFASV FTAASIVWQL GEPKLAMPFV LGIIAGGLVD
51 LDNXXTGRLK NIITTVLFT LSSLTAQSTL GTGLPFILAM TLMTXXTIL
101 GAX...
```

Further work revealed the complete nucleotide sequence <SEQ ID 105>:

```
15 1 ATGAAAACCC CACTCCTCAA GCCTCTGCTC ATTACCTCGC TTCCCCTTTT
51 CGCCAGTGTT TTTACCGCCG CCTCCATCGT CTGGCAGCTA GGCGAACCCA
101 AGCTCGCCAT GCCCTTCGTA CTCGGCATCA TCGCCGGCGG CTTTGTGGAT
151 TTGGACAACC GCCTGACCGG ACGGCTGAAA AACATCATCA CCACCGTCGC
201 CCTGTTCAACC CTCTCCTCGC TCACGGCACA AAGCACCTC GGCACAGGGC
20 251 TGCCCTTCAT CCTCGCCATG ACCCTGATGA CTTTGGCTT CACCATTTTA
301 GGCGCGGTCG GGCTCAAATA CCGCACCTTC GCCTTCGGTG CACTCGCCGT
351 CGCCACCTAC ACCACACTTA CCTACACCCC CGAAACCTAC TGGCTGACCA
401 ACCCTTCAT GATTTTATGC GGCACCGTAC TGTACAGCAC CGCCATCCTC
451 CTGTTCCAAA TCGTCCCTGCC CCACCGCCCC GTCCAAGAAA GCGTCGCCAA
25 501 CGCCTACGAC GCACTCGGCG GCTACCTCGA AGCCAAAGCC GACTTCTTCG
551 ACCCCGATGA GGCAGCCTGG ATAGGCAACC GCCACATCGA CCTCGCCATG
601 AGCAACACCG GCGTCATCAC CGCTTCAAC CAATGCCGTT CCGCCCTGTT
651 TTACCGCCTT CGCGGCAAAC ACCGCCACCC GCGCACCGCC AAAATGCTGC
701 GTTACTACTT TGCCGCCCAA GACATACAG AACGCATCAG CTCCGCCAC
30 751 GTCGATTATC AGGAAATGTC CGAAAAATTC AAAAACACCG ACATCATCTT
801 CGGCATCCAC CGCCTGCTCG AAATGCAGGG ACAAGCCTGC CGCAACACCG
851 CCCAAGCCCT GCGCGCAAGC AAAGACTACG TTTACAGCAA ACGCCTCGGC
901 CGCGCCATCG AAGGCTGCCG CCAATCGCTG CGCCTCCTTT CAGACAGCAA
951 CGACAGTCCC GACATCCGCC ACCTGCGCGC CTTTCTCGAC AACCTCGGCA
35 1001 GCGTCGACCA GCAGTTCCGC CAACTCCAGC ACAACGGCCT GCAGGCAGAA
1051 AACGACCGCA TGGGCGACAC CCGCATCGCC GCCCTCGAAA CCAGCAGCCT
1101 CAAAAACACC TGGCAGGCAA TCCGTCCGCA GCTAAACCTC GAATCAGGCG
1151 TATTCCGCCA TGCCGTCCGC CTGTCCCTCG TCGTTGCCGC CGCCTGCACC
40 1201 ATCGTCGAAG CCCTCAACCT CAACCTCGGC TACTGGATAC TACTGACCGC
1251 CCTTTTCGTC TGCCAACCCA ACTACACCGC CACCAAAAGC CGCGTCCGCC
1301 AGCGCATCGC CGGCACCGTA CTCGGCGTAA TCGTCGGCTC GCTCGTCCCC
1351 TACTTCACCC CGTCTGTCGA AACCAACTC TGGATTGTCA TCGCCAGTAC
1401 CACCCTCTTT TTCATGACCC GCACCTACAA ATACAGTTT TCCACCTTCT
45 1451 TCATTACCAT TCAAGCCCTG ACCAGCCTCT CCCTCGCAGG TTTGGACGTA
1501 TAGCGCGCCA TGCCGTACG CATCATCGAC ACCATTATCG GCGCATCCCT
1551 TGCTTGGCGC GCAGTCAGCT ACCTGTGGCC AGACTGGAAA TACCTCACGC
1601 TCGAACGCAC CGCCGCCCTT GCCGTATGCA GCAACGGTGC CTATCTCGAA
1651 AAAATCACCG AACGCCTCAA AAGCGGCGAA ACCGGCGACG ACGTCGAATA
50 1701 CGCGGCCACC CGCCGCCGCG CCCACGAACA CACCGCCGCC CTCAGCAGCA
1751 CCCTTTCCGA CATGAGCAGC GAACCGCAA AATTCGCCGA CAGCCTGCAA
1801 CCGGGCTTTA CCCTGCTCAA AACCGGTAC GCCCTGACCG GCTACATCTC
1851 CGCCCTCGCG GCATACCGCA GCGAAATGCA CGAAGAATGC AGCCCCGACT
1901 TTACCGCACA GTTCCACCTC GCCGCCGAAC ACACCGCCA CATCTTCCAA
1951 CACCTGCCCG AAACCGAACC CGACGACTTT CAGACAGCAC TGGATACACT
55 2001 GCGCGGCGAA CTCGACACCC TCCGCACCCA CAGCAGCGGA ACACAAAGCC
2051 ACATCCTCCT CCAACAGCTC CAATCATCG CCCGACAGCT CGAACCCTAC
2101 TACCGCGCCT ACCGCCAAAT TCCGCACAGG CAGCCCCAAA ATGCAGCCTG
2151 A
```

This corresponds to the amino acid sequence <SEQ ID 106; ORF19-1>:

```
60 1 MKTPLLKPLL ITSLPVFASV FTAASIVWQL GEPKLAMPFV LGIIAGGLVD
```

51 LDNRLTGRLE NIITVALFT LSSLTAQSTL GTGLPFILAM TLMTEGFTIL  
 101 GAVGLKYRTF AFGALAVATY TTLTYTPETY WLTNPFMILC GTVLYSTAIL  
 151 LFQIVLPHRP VQESVANAYD ALGGYLEAKA DFFDPDEAAW TGNRHIDLAM  
 201 SNTGVITAFN QCRSALFYRL RGKHRHPTA KMLRYFAAQ DIHERISSAH  
 251 VDYQEMSEKF KNTDIIFRIH RLLEMQGQAC RNTAQALRAS KDYVYSKRLG  
 301 RAIEGCRQSL RLLSDSNDSP DIRHLRRLD NLGSVDQQFR QLQHNGLQAE  
 351 NDRMGDTRIA ALETSSLKNT WQAIRPQLNL ESGVFRHAVR LSLVVAAACT  
 401 IVEALNLLG YWILLTALFV CQPNYTATKS RVRQRIAGTV LGVIVGSLVP  
 451 YFTPSVETKL WIVIASTTLF FMTRTYKYSF STFFITIQAL TSLSLAGLDV  
 501 YAAMPVRIID TIIGASLAWA AVSYLWPDWK YLTLETAAL AVCSNGAYLE  
 551 KITERLKSGE TGDDVEYRAT RRAHEHTAA LSSTLSDMSS EPAKFADSLQ  
 601 PGFTLLKTGY ALTGYISALG AYRSEMHEEC SPDFTAQFHL AAEHTAHIFQ  
 651 HLPETEPDDF QTALDTLRGE LDTLRTHSSG TQSHILLQQL QLIARQLEPY  
 701 YRAYRQIPHR QPQNAA\*

15 Computer analysis of this amino acid sequence gave the following results:

Homology with predicted transmembrane protein YHFK of *H. influenzae* (accession number P44289)

ORF19 and YHFK proteins show 45% aa identity in 97 aa overlap:

20 orf19 6 LKPLLITSLPVFASVFTAASIVWQLGEPKLPFVLGIIAGGLVDLDNXXTGRLEKNIITT 65  
 L +I+++PVF +V AA +W +MP +LGIIAGGLVDLDN TGRLEK+ T  
 YHFK 5 LNAKIVISTIPVFIADVNTAAVGIWFFDISSQSMPLILGIIAGGLVDLDNRLTGRLEKNVFFT 64  
 25 orf19 66 VALFTLSSLTAQSTLGTGLPFILAMTLMXXFTILGA 102  
 + F++SS Q +G + +I+ MT++T FT++GA  
 YHFK 65 LIAFSSISFIVQLHIGKPIQYIVLMTVLTFTFTMIGA 101

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF19 shows 92.2% identity over a 102aa overlap with an ORF (ORF19a) from strain A of *N. meningitidis*:

30 orf19.pep 10 20 30 40 50 60  
 MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLPFVLGIIAGGLVDLDNXXTGRLEK  
 orf19a 10 20 30 40 50 60  
 MKTPPLKPLLITSLPVFASVFTAASIVWQLGEPKLPFVLGIIAGGLVDLDNRLTGRLEK  
 35 orf19.pep 70 80 90 100  
 NIITVALFTLSSLTAQSTLGTGLPFILAMTLMXXFTILGAX  
 orf19a 70 80 90 100 110 120  
 NIIATVALFTLSSLVAQSTLGTGLPFILAMTLMTEGFTIMGAVGLKYRTFAFGALAVATY  
 40 orf19a 130 140 150 160 170 180  
 TTLTYTPETYWLTNPFMILCGTVLYSTAILFQIILPHRPVQENVANAYEALGSYLEAKA

The complete length ORF19a nucleotide sequence <SEQ ID 107> is:

45 1 ATGAAAACCC CACCCCTCAA GCCTCTGCTC ATTACCTCGC TTCCCGTTTT  
 51 CGCCAGTGTC TTTACCGCCG CCTCCATCGT CTGGCAGCTG GGCGAACCCA  
 101 AGCTCGCCAT GCCCTTCGTA CTCGGCATCA TCGCTGGCGG CCTGGTCGAT  
 151 TTGGACAACC GCCTGACCGG ACGGCTGAAA AACATCATCG CCACCGTCGC  
 201 CCTGTTCCAC CTCTCCTCAC TTGTCGCGCA AAGCACCCCTC GGCACAGGTT  
 251 TGCCATTCAT CCTCGCCATG ACCCTGATGA CTTTCGGCTT TACCATCATG  
 301 GGCGCGGTCT GGCTGAAATA CCGCACCTTC GCCTTCGGCG CACTCGCCGT  
 351 CGCCACCTAC ACCACACTTA CCTACACCCC CGAAACCTAC TGCTGACCA  
 401 ACCCCTTTAT GATTCTGTGC GGAACCGTAC TGTACAGCAC CGCCATCATC  
 451 CTGTTCCAAA TCATCCTGCC CCACCGCCCC GTTCAAGAAA ACGTCGCCAA  
 501 CGCCTACGAA GCACTCGGCA GCTACCTCGA AGCCAAAGCC GACTTTTTTCG  
 551 ATCCCGACGA AGCCGAATGG ATAGGCAACC GCCACATCGA CCTCGCCATG  
 601 AGCAACACCG GCGTCATCAC CGCCTTCAAC CAATGCCGTT CCGCCCTGTT  
 651 TTACCGCCTT CGCGGCAAAAC ACCGCCACCC GCGCACCGCC AAAATGCTGC  
 701 GCTACTACTT CGCCGCCCAA GACATACAG AACGCATCAG CTCCGCCAC  
 751 GTCGACTACC AAGAGATGTC CGAAAAATTC AAAAACACCG ACATCATCTT



5

10

15

20

25

30

35

40

50

55

60

65

	orf19-1	DDFDPDEAAWIGNRHIDLAMSNTGVITAFNQCRSALFYRLRGKHRHPRTAKMLRYYFAAQ	190	200	210	220	230	240
5	orf19a.pep	DIHERISSAHVDYQEMSEKFKNTDIIIFRIHRLLEMQGOACRNTAQALRASKDYVYSKRLG	250	260	270	280	290	300
	orf19-1	DIHERISSAHVDYQEMSEKFKNTDIIIFRIHRLLEMQGOACRNTAQALRASKDYVYSKRLG	250	260	270	280	290	300
10	orf19a.pep	RAIEGCRQSLRLSDSNDNPDIRHLRRLDNLGSVDQQFRQLQHNGLQAENDRMGDTRIA	310	320	330	340	350	360
	orf19-1	RAIEGCRQSLRLSDSNDSPDIRHLRRLDNLGSVDQQFRQLQHNGLQAENDRMGDTRIA	310	320	330	340	350	360
15	orf19a.pep	ALETGSLKNTWQAIRPQINLESGVFRHAVRLSLVVAACTIVEALNINLGYWILLTALFV	370	380	390	400	410	420
20	orf19-1	ALETSSLKNTWQAIRPQINLESGVFRHAVRLSLVVAACTIVEALNINLGYWILLTALFV	370	380	390	400	410	420
25	orf19a.pep	CQPNYTATKSRVRQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIASTTLFFMTRTYKYSF	430	440	450	460	470	480
	orf19-1	CQPNYTATKSRVRQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIASTTLFFMTRTYKYSF	430	440	450	460	470	480
30	orf19a.pep	STFFITIQAALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTLETAAL	490	500	510	520	530	540
	orf19-1	STFFITIQAALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTLETAAL	490	500	510	520	530	540
35	orf19a.pep	AVCSNGAYLEKITERLKSGETGDDVEYRATRRRAHEHTAALSSTLSDMSSEPAKFADSLQ	550	560	570	580	590	600
	orf19-1	AVCSNGAYLEKITERLKSGETGDDVEYRATRRRAHEHTAALSSTLSDMSSEPAKFADSLQ	550	560	570	580	590	600
40	orf19a.pep	PGFTLLKTGYALTGYISALGAYRSEMHEECSPDFTAQFHAAEHTAHIFQHLPETEPDDF	610	620	630	640	650	660
	orf19-1	PGFTLLKTGYALTGYISALGAYRSEMHEECSPDFTAQFHAAEHTAHIFQHLPETEPDDF	610	620	630	640	650	660
45	orf19a.pep	QTALDTLRGELDTLRTHSSGTQSHILLQQLLIARQLEPYRAYRQIPHRQPQNAAX	670	680	690	700	710	
50	orf19-1	QTALDTLRGELDTLRTHSSGTQSHILLQQLLIARQLEPYRAYRQIPHRQPQNAAX	670	680	690	700	710	

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF19 shows 95.1% identity over a 102aa overlap with a predicted ORF (ORF19.ng) from *N.gonorrhoeae*:

55	orf19.pep	MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLPMPFVLGIIAGGLVDLDNXXTGRLK	60
	orf19ng	MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLPMPFVLGIIAGGLVDLDNRLTGRLK	60
60	orf19.pep	NIITTVALLFTLSSSLTAQSTLTGTLPPFILAMTLMTXXFTILGAX	103
	orf19ng	NIIATVALLFTLSSSLTAQSTLTGTLPPFILAMTLMTFGTILGAVGLKYRTAFAGALAVATY	120

An ORF19ng nucleotide sequence <SEQ ID 109> is predicted to encode a protein having amino acid sequence <SEQ ID 110>:

5  
1 MKTPLLKPLL ITSLPVFASV FTAASIVWQL GEPKLAMPFV LGIIAGGLVD  
51 LDNRLTGRLK NIIATVALFT LSSLTAQSTL GTGLPFILAM TLMTFGFTIL  
101 GAVGLKYRTF AFGALAVATY TTLTYTPETY WLTNPFMILC GTVLYSTAI I  
151 LFQIILPHRP VQESVANAYE ALGGYLEAKA DFFDPDEAAW IGNRHIDLAM  
201 SNTGVITAFN QCRSALFYRL RGKRRHPRTA KMLRYYFAAQ DIHERISSAH  
251 VDYQEMSEKF KNTDIIFRIR RLEMQGQAC RNTAQAIRSG KDYVYSKRLG  
301 RAIEGCRQSL RLLSDGNDSP DIRHLSRLLD NLGSVDQQFR QLRHSDSPA E  
351 NDRMGDTRIA ALETGSFKNT \*

Further work revealed the complete nucleotide sequence <SEQ ID 111>:

10  
1 ATGAAACCC CACTCCTCAA GCCTCTGCTC ATTACCTCGC TTCCCGTTTT  
51 CGCCAGTGTC TTTACCGCCG CCTCCATCGT CTGGCAGCTA GCGGAACCCA  
101 AGCTCGCCAT GCCCTTCGTA CTCGGCATCA TCGCCGGCGG CCTGGTCGAT  
151 TTGGACAACC GCCTGACCGG ACGGCTGAAA AACATCATCG CCACCGTCGC  
15  
201 CCTGTTTACC CTCTCCTCGC TCACGGCGCA AAGCACCTC GGCACAGGGC  
251 TGCCCTTCAT CCTCGCCATG ACCCTGATGA CCTTCGGCTT TACCATTTTA  
301 GGCGCGGTGCG GGCTGAAATA CCGCACCTTC GCCTTCGGCG CACTCGCCGT  
351 CGCCACCTAC ACCACGCTTA CCTACACCCC CGAAACCTAC TGGCTGACCA  
401 ACCCCTTCAT GATTTTATGC GGCACCGTAC GTTACAGCAC CGCCATCATC  
451 CTGTTCCAAA TCATCCTGCC CCACCGCCCC GTCCAAGAAA GCGTCGCCAA  
20  
501 TGCTACGAA GCACTCGGCG GCTACCTCGA AGCCAAAGCC GACTTCTTCG  
551 ACCCGGATGA GGCAGCCTGG ATAGGCAACC GCCACATCGA CCTCGCCATG  
601 AGCAACACCG GCGTCATCAC CGCCTTCAAC CAATGCCGTT CCGCCCTGTT  
651 TTACCGTTTG CGCGGCAAAAC ACCGCCACCC GCGCACCGCC AAAATGCTGC  
25  
701 GCTACTACTT CGCCGCCCAA GACATCCACG AACGCATCAG CTCGCCAC  
751 GTCGACTACC AAGAGATGTC CGAAAAATTC AAAACACCG ACATCATCTT  
801 CGGCATCCGC CGCTGCTCG AAATGCAGGG GCAGGCGTGC CGCAACACCG  
851 CCAAGCCAT CCGTTCGGCG AAAGACTAcg tTTACAGCAA ACGCCTCGGA  
901 CGCGCCATcg aaggctgCCG CCAGTCGCTg cgcctCCTt cagacggcaA  
951 CGACAGTCCC GACATCCGCC ACCTGAGccg CCTTCTCGAC AACCTCGgca  
30  
1001 GCGTcgacca gcagtTCgc caactCCGAC ACAGcgactC CCCCgcgaa  
1051 Aacgaccgca tggcgacaca CCGCATCGCC GCCctcgaaa cggcgagctT  
1101 caaaaaCAcc tggcaggCAA TCCGTCCGCa gctgaaCCTC GAATCatgCG  
1151 TATTCCGCCA TGCCGTCCGC CTGTCCCTCG TCGTTGCCGC CGCCTGCACC  
1201 ATCGTCgaag cCCTCAACCT CAACCTCGGC TACTGGATAC TGCTGACCCG  
35  
1251 CCTTTTCGTC TGCCAACCCA ACTACACCGC CACCAAAAGC CGCGTGTACC  
1301 AACGCATCGC CGGCACCGTA CTCGGCGTAA TCGTCGGCTC GCTCGTCCCC  
1351 TACTTCACCC CCTCCGTCGA AACCAAATC TGGATTGTCA TCGCCGGTAC  
1401 CACCTGTTC TTCAATGACC GCACCTACAA ATACAGTTTC TCCACCTTCT  
40  
1451 TCATCACCAT TCAGGCACTG ACCAGCCTCT CCCTCGCAGG TTTGGACGTA  
1501 TACGCCGCCA TGCCCGTGC GATCATcgac ACCATTATCG GCGCATCCCT  
1551 TGCTGGGCG CGGTCACTG ACCTGTGGCC AGACTGGAAA TACCTCACGC  
1601 TCGAACGCAC CGCCGCCCTT GCCGTATGCA GCAGCGGCAC ATACCTCCAA  
1651 AAAATTGCCG AACGCCTCAA AACCGGCGAA ACCGGCGAGC ACATAGAATA  
45  
1701 CCGCATCACC CGCCGCCGCG CCCACGAACA CACCGCCGCC CTCAGCAGCA  
1751 CCCTTTCCGA CATGAGAGC GAACCCGCAA AATTCCGCCA CAGCCTGCAA  
1801 CCGGCTTTA CCCTGCTCAA AACCGGTAC GCCCTGACCG GCTACATCTC  
1851 CGCCCTCGGC GCATACCGCA GCGAAATGCA CGAAGAATGC AGCCCCGACT  
1901 TTACCGCACA GTTCCACCTT GCGCCGAAC ACACGCCCA CATCTTCCAA  
50  
1951 CACCTGCCCG ACATGGGACC CGACGACTTT CAGACGGCAT TGGATACAGT  
2001 GCGCGGCGAA CTCGGCACCC TCCGCACCG CAGCAGCGGA ACACAAAGCC  
2051 ACATCTCCT CCAACAGCTC CAACTCATCG CccgGCAACT CGAACCTTAC  
2101 TACCGGCCT ACCGACAAAT TCCGCACAGG CAGCCCCAAA ACGCAGCCTG  
2151 A

This corresponds to the amino acid sequence <SEQ ID 112; ORF19ng-1>:

55  
1 MKTPLLKPLL ITSLPVFASV FTAASIVWQL GEPKLAMPFV LGIIAGGLVD  
51 LDNRLTGRLK NIIATVALFT LSSLTAQSTL GTGLPFILAM TLMTFGFTIL  
101 GAVGLKYRTF AFGALAVATY TTLTYTPETY WLTNPFMILC GTVLYSTAI I  
151 LFQIILPHRP VQESVANAYE ALGGYLEAKA DFFDPDEAAW IGNRHIDLAM  
60  
201 SNTGVITAFN QCRSALFYRL RGKRRHPRTA KMLRYYFAAQ DIHERISSAH  
251 VDYQEMSEKF KNTDIIFRIR RLEMQGQAC RNTAQAIRSG KDYVYSKRLG  
301 RAIEGCRQSL RLLSDGNDSP DIRHLSRLLD NLGSVDQQFR QLRHSDSPA E  
351 NDRMGDTRIA ALETGSFKNT WQAIRPQLNL ESCVFRHAVR LSLVVAACCT  
401 IVEALNINLG YWILLTALFV CQPNYTATKS RVYQRIAGTV LGVIVGSLVP  
65  
451 YFTPSVETKL WIVIAGTTLF FMTRTYKYSF STFFITIQL TSLSLAGLDV  
501 YAAMPVRIID TIIGASLAWA AVSYLWPDWK YLTLERTAAL AVCSSGTYLQ  
551 KIAERLKTGE TGDDIEYRIT RRAHEHTAA LSSTLSDMSS EPAKFADSLQ

601 PGFTLLKTGY ALTGYISALG AYRSEMHEEC SPDFTAQFHL AAEHTAHIFQ  
651 HLPDMGPDDF QTALDTLRGE LGTLRTRSSG TQSHILLQOL QLIARQLEPY  
701 YRAYRQIPHR QPQNAA\*

ORF19ng-1 and ORF19-1 show 95.5% identity in 716 aa overlap:

5  
orf19-1.pep 10 20 30 40 50 60  
MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNRLTGRLK  
|||  
orf19ng-1 10 20 30 40 50 60  
MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNRLTGRLK  
10  
orf19-1.pep 70 80 90 100 110 120  
NIIITVALFTLSSLTAQSTLGTGLPFILAMTLMTFGFTILGAVGLKYRTFAFGALAVATY  
|||:|  
orf19ng-1 70 80 90 100 110 120  
NIIATVALFTLSSLTAQSTLGTGLPFILAMTLMTFGFTILGAVGLKYRTFAFGALAVATY  
15  
orf19-1.pep 130 140 150 160 170 180  
TTLTYTPETYWLTNPFMILCGTVLYSTAILLFQIVLPHRPVQESVANAYDALGGYLEAKA  
|||  
orf19ng-1 130 140 150 160 170 180  
TTLTYTPETYWLTNPFMILCGTVLYSTAILLFQIILPHRPVQESVANAYEALGGYLEAKA  
20  
orf19-1.pep 190 200 210 220 230 240  
DFFDPDEAAWIGNRHIDLAMSGNTGVITAFNQCRSALFYRLRGKHRHPRTAKMLRYYFAAQ  
|||  
orf19ng-1 190 200 210 220 230 240  
DFFDPDEAAWIGNRHIDLAMSGNTGVITAFNQCRSALFYRLRGKHRHPRTAKMLRYYFAAQ  
25  
orf19-1.pep 250 260 270 280 290 300  
DIHERISSAHVDYQEMSEKFNKTDIIFRIHRLLEMQGOACRNTAQALRASGDYVYSKRLG  
|||:|  
orf19ng-1 250 260 270 280 290 300  
DIHERISSAHVDYQEMSEKFNKTDIIFRIHRLLEMQGOACRNTAQAIRSGKDYVYSKRLG  
30  
orf19-1.pep 310 320 330 340 350 360  
RAIEGCRQSLRLLSDSNDSPDIRHLRRLLDNLGSVDQQFRLQHNGLQENDRMGDTRIA  
|||:|  
orf19ng-1 310 320 330 340 350 360  
RAIEGCRQSLRLLSDGNDSPDIRHLSRLLDNLGSVDQQFRLRHSDSPAENDRMGDTRIA  
40  
orf19-1.pep 370 380 390 400 410 420  
ALETSSLKNTWQAIKPNLNLSEGVFRHAVRLSLVVAACTIVEALNLLNGYWILLTALFV  
|||:|  
orf19ng-1 370 380 390 400 410 420  
ALETGSFKNTWQAIKPNLNLSEGVFRHAVRLSLVVAACTIVEALNLLNGYWILLTALFV  
45  
orf19-1.pep 430 440 450 460 470 480  
CQPNYTATKSRVRQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIASTTLFEMTRTYKYSF  
|||  
orf19ng-1 430 440 450 460 470 480  
CQPNYTATKSRVYQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIASTTLFEMTRTYKYSF  
50  
orf19-1.pep 490 500 510 520 530 540  
STFFITIQALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTTLERTAAAL  
|||  
orf19ng-1 490 500 510 520 530 540  
STFFITIQALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTTLERTAAAL  
55  
orf19-1.pep 550 560 570 580 590 600  
AVCSNGAYLEKITERLKSGETGDDVEYRATRRRAHEHTAALSSTLSDMSSEPAKFADSLQ  
|||:|  
orf19ng-1 550 560 570 580 590 600  
AVCSSGYLQKIAERLKTGETGDDIEYRITRRRAHEHTAALSSTLSDMSSEPAKFADSLQ  
60  
orf19-1.pep 610 620 630 640 650 660  
PGFTLLKTGYALTGYISALGAYRSEMHEECSPDFTAQFHLAAEHTAHIFQHLPETEPDDF  
|||  
orf19ng-1 610 620 630 640 650 660  
PGFTLLKTGYALTGYISALGAYRSEMHEECSPDFTAQFHLAAEHTAHIFQHLPETEPDDF

		610	620	630	640	650	660
		670	680	690	700	710	
5	orf19-1.pep	QTALD	TLRGEL	DLRTH	SSGTQ	SHILL	QQLQLIARQLEPYRAYRQIPHROPQNAAX
	orf19ng-1	QTALD	TLRGEL	GLTR	SSGTQ	SHILL	QQLQLIARQLEPYRAYRQIPHROPQNAAX
		670	680	690	700	710	

In addition, ORF19ng-1 shows significant homology to a hypothetical gonococcal protein previously entered in the databases:

10	sp 033369 YOR2_NEIGO HYPOTHETICAL 45.5 KD PROTEIN (ORF2) gnl PID e1154438 (AJ002423) hypothetical protein [Neisseria gonorrh] Length = 417 Score = 1512 (705.6 bits), Expect = 5.3e-203, P = 5.3e-203 Identities = 301/326 (92%), Positives = 306/326 (93%)
15	Query: 307 RQSLRLSDGNDSPDIRHLSRLLDNLGSVDQQFRQLRHSDSPAENDRMGDTRIAALETGS 366 RQSLRLSDGNDSDIRHLSRLLDNLGSVDQQFRQLRHSDSPAENDRMGDTRIAALETGS Sbjct: 1 RQSLRLSDGNDSDIRHLSRLLDNLGSVDQQFRQLRHSDSPAENDRMGDTRIAALETGS 60
20	Query: 367 FKNTWQAIROPQNLNLESCVFRHAVRLSLVVAACTIVEALNINLGYWILLTALFVCQPNYT 426 FKNTWQAIROPQNLNLESCVFRHAVRLSLVVAACTIVEALNINLGYWILLTALFVCQPNYT Sbjct: 61 FKNTWQAIROPQNLNLESCVFRHAVRLSLVVAACTIVEALNINLGYWILLTALFVCQPNYT 120
25	Query: 427 ATKSRVYQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIAGTTLFMTRTYKYSFSTFFIT 486 ATKSRVYQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIAGTTLFMTRTYKYSFSTFFIT Sbjct: 121 ATKSRVYQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIAGTTLFMTRTYKYSFSTFFIT 180
30	Query: 487 IQALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTTLERTAALAVCSSG 546 IQALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTTLERTAALAVCSSG Sbjct: 181 IQALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTTLERTAALAVCSSG 240
35	Query: 547 TYLQKIAERLKTGETGDDIEYRITRRRAHEHTAALSSTLSDMSSEPAKFADSLQPGFTLL 606 TYLQKIAERLKTGETGDDIEYRITRRRAHEHTAALSSTLSDMSSEPAKFADSLQPGFTLL P Sbjct: 241 TYLQKIAERLKTGETGDDIEYRITRRRAHEHTAALSSTLSDMSSEPAKFADTCNPALPCS 300
	Query: 607 KTG YALTGYISALGAYRSEMHEECSP 632 K ALTGYISALG ++ + +P Sbjct: 301 KPATALTGYISALGHTAAKCTKNAAP 326

Based on this analysis, including the presence of several putative transmembrane domains in the gonococcal protein (the first of which is also seen in the meningococcal protein), and on homology with the YHFK protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 14

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 113>:

45	1	ATGAATATGC	TGGGAGCTTT	GGCAAAAGTC	GGCAGCCTGA	CGATGGTGTG
	51	GCGCGTTTTG	GGATTTGTGC	GCGATACGGT	CATTGCGCGG	GCATTCGGCG
	101	CGGGTATGGC	GACGGATGCG	TTTTTTGTGC	CGTTCAAACT	GCCCAACCTG
	151	CTTCGCGCGC	TGTTTTCGGA	GGGGGCGTTT	GCCCAAGCGT	TTGTGCCGAT
50	201	TTTGGCGGAA	TACAAGGAAA	CGCGTTCAAA	AGAGGCGG.C	GAAGCCTTTA
	251	TCCGCCATGT	GGCGGGGATG	CTGTCGTTTG	TACTGGTTAT	CGTTACCGCG
	301	CTGGGCATAC	TTGCCGCGCC	TTGGGTGATT	TATGTTTCCG	CACCCGAGTT
	351	TTGCCCAAGA	TGCCGACAAA	TTTCAGCTCT	CCATCGATT	GCTGCGGATT
	401	ACGTTTCCTT	ATATATTATT	GATTTCCCTG	TCTTCATTG	TCGGCTCGGT
55	451	ACTCAATTCT	TATCATAAGT	TCGGCATTCC	GGCGTTTACG	CCAC.GTTTC
	501	TGAACGTGTC	GTTTATCGTA	TTCCGCGTGT	TTTTCGTGCC	GTATTTTCGAT

551 CCGCCCGTTA CCGCGCyGGC GTGGGCGGTC TTTGTCGGCG GCATTTTGA  
 601 ACTCGmTTC CAACTGCCCT GGCTGGCGAA ACTGGGCTTT TTGAAACTGC  
 651 CCAAACtGAG TTTCAAAGAT GCGGCGGTCA ACCGCGTGAT GAAACAGATG  
 701 GCGCCTGCgA TTTTgGGCGT GAqCGTGGCG CAGGTTTCTT TGGTGATCAA  
 751 CACGATTTTc GCGTCTTATC TGCAATCGGG CAGCGTTTCA TGGATGTATT  
 801 ACGCCGACCG CATGATGGAG CTGCCCAGCG GCGTGCTGGG GCGGGCACTC  
 851 GGTACGATTT TGCTGCCGAC TTTGTCCAAA CACTCGGCAA ACCaAGATAC  
 901 GGaACAGTTT TCCGCCCTGC TCGACTGGGG TTTGCGCCTG TGCATGctgc  
 951 TGACGCTGCC GGCGgcGGTC GGACTGGCGG TGTGTCTGTT cCCgCtGGTG  
 1001 GCGACGCTGT TTATGTACCG CGwATTTACG CTGTTTGACG CGCAGATGAC  
 1051 GCAACACGCG CTGATTGCCT ATTCTTTTCGG TTAATCGGC TTAATCATGA  
 1101 TTAAAGTGTT GGCACCCGGC TTCTATGCGC GGCAAACAT CAAwAmCCCC  
 1151 GTCAAATCG CCATCTTCAC GCTCATCTGC mCGCAGTTGA TGAACCTTGs  
 1201 CTTTAYCGGC CCACTrrAAC rCasTCGGAC TTTTCGCTGC CATCGGTCTG  
 1251 GGCGCGTGTA TCAATGCCGG ATTGTTGTTT TACCTGTTGC GCAGACACGG  
 1301 TATTTACCAA CCTGG.CAAG GGTGGGCAG CGTTCTT.AG CAAAATGCT  
 1351 GcTCTCGCTC GCCGTGA

This corresponds to the amino acid sequence <SEQ ID 114; ORF20>:

1 MNMLGALAKV GSLTMVSRVL GFVRDTVIAF AFGAGMATDA FFFVAFKLPNL  
 51 LRRVFAEGAF AQAFVPILAE YKETRSKEAX EAFIRHVAGM LSFVLVIVTA  
 101 LGILAAPWVI YYSAPSFQAD ADKFLSIDL LRITFPYILL ISLSSFVGSV  
 151 LNSYHKFGIP AFTFXFLNVS FIVFALFFVP YFDPFVTAXA WAVFVGGILQ  
 201 LXFQLPWLA LGLKLPKLS FKDAAVNRVM KQMAPAILGV SVAQVSLVIN  
 251 TIFASYLQSG SVSWMYADY MMELPSGVLG AALGTILLPT LSKHSANQDT  
 301 EQFSALLDWG LRLCMLLTLP AAVGLAVLSF PLVATLFMYR XFTLFDAQMT  
 351 QHALIAYSFG LIGLIMIKVL APGFYARQNI XXPVKIAIFT LICXQLMNLX  
 401 FXGPLXXIGL SLAIGLGACI NAGLLFYLLR RHGIYQXPQG LGSVLXQKCC  
 451 SRSP\*

These sequences were elaborated, and the complete DNA sequence <SEQ ID 115> is:

1 ATGAATATGC TGGGAGCTTT GGCAAAAGTC GGCAGCCTGA CGATGGTGTC  
 51 GCGCGTTTTG GGATTTGTGC GCGATACGGT CATTCGCGCG GCATTCGGCG  
 101 CCGGTATGGC GACGGATGCG TTTTGTGTCG CGTTCAAAC GCGCAACCTG  
 151 CTTCGCCGCG TGTTCGCGA GGGGCGTTT GCCCAAGCGT TTGTGCCGAT  
 201 TTTGGCGGAA TACAAGGAAA CGCGTTCAA AGAGGCGGCG GAGGCTTTTA  
 251 TCCGCCATGT GCGGGGATG CTGTCGTTTG TACTGGTTAT CGTTACCGCG  
 301 CTGGGCATAC TTGCCGCGCC TTGGGTGATT TATGTTCCG CACCCGGTTT  
 351 TGCCCAAGAT GCCGACAAAT TTCAGCTCTC CATCGATTG CTGCGGATTA  
 401 CGTTTCCTTA TATATTATG ATTCCCTGT CTTCATTTGT CGGCTCGGTA  
 451 CTCAATCTT ATCATAAGT CGGCATTCCG GCGTTTACGC CCACGTTTCT  
 501 GAACGTGTCG TTTATCGTAT TCGCGCTGTT TTTGTCGCCG TATTTCGATC  
 551 CGCCCGTTAC CGCGCTGGCG TGGGCGGTCT TTGTCGGCGG CATTTTGCAA  
 601 CTCGGCTTCC AACTGCCCTG GCTGGCGAAA CTGGGCTTTT TGAACCTGCC  
 651 CAAACTGAGT TTCAAAGATG CGGCGGTCAA CCGCGTGATG AAACAGATGG  
 701 CGCCTGCGAT TTGGGCGGTG AGCGTGCGCG AGGTTTCTTT GGTGATCAAC  
 751 ACGATTTTCG CGTCTTATCT GCAATCGGGC AGCGTTTCAT GGATGTATTA  
 801 CGCCGACGCG ATGATGGAGC TGCCAGCGG CGTGCTGGGG GCGGCACTCG  
 851 GTACGATTTT GCTGCCGACT TTGTCCAAAC ACTCGGCAA CCAAGATACG  
 901 GAACAGTTT CCGCCCTGCT CGACTGGGGT TTGCGCCTGT GCATGCTGCT  
 951 GACGCTGCCG GCGGCGGTG GACTGGCGGT GTTGTCTGTC CCGCTGGTGG  
 1001 CGACGCTGTT TATGTACCGC GAATTTACGC TGTGACGC GCAGATGACG  
 1051 CAACACGCGC TGATTGCCCTA TTCTTTCGGT TTAATCGGCT TAATCATGAT  
 1101 TAAAGTGTTG GCACCCGGCT TCTATGCGCG GCAAAACATC AAAACGCCCG  
 1151 TCAAAATCGC CATCTTCACG CTCATCTGCA CGCAGTTGAT GAACCTTGCC  
 1201 TTTATCGGCC CACTGAAACA CGTCGACTT TCGCTTGCCA TCGGTCTGGG  
 1251 CGCGTGATC AATGCCGGAT TGTGTTTGA CCTGTTGCGC AGACACGGTA  
 1301 TTTACCAACC TGGCAAGGGT TGGCAGCGT TCTAGCAAA AATGCTGCTC  
 1351 TCGCTCGCCG TGATGTGCGG CGGACTGTGG GCAGCGCAGG CTTACCTGCC  
 1401 GTTTGAATGG GCGCACGCGC GCGGAATGCG GAAAGCGGG CAGCTCTGCA  
 1451 TCCTGATTGC CGTCGGCGGC GGACTGTATT TCGCATCACT GCGGCTTTG  
 1501 GGCTTCCGTC CCGCCATTT CAAACGCGTG GAAACTGA

This corresponds to the amino acid sequence <SEQ ID 116; ORF20-1>:

1 MNMLGALAKV GSLTMVSRVL GFVRDTVIAF AFGAGMATDA FFFVAFKLPNL  
 51 LRRVFAEGAF AQAFVPILAE YKETRSKEAA EAFIRHVAGM LSFVLVIVTA

101 LGILAAPWVI YVSAPGFAQD ADKFQLSIDL LRITFPYILL ISLSSEFVGSV  
 151 LNSYHKFGIP AFTPTFLNVS FIVFALFFVP YFDPPTALA WAVFVGGILO  
 201 LGFQLPWLAK LGFLKLPKLS FKDAAVNRVM KQMAPAILGV SVAQVSLVIN  
 251 TIFASYLQSG SVSWMYADR MMELPSGVLG AALGTILLPT LSKHSANQDT  
 301 EQFSALLDWG LRLCMLLTLP AAVGLAVLSF PLVATLFMYR EFTLFDAQMT  
 351 QHALIAYSFG LIGLIMIKVL APGFYARQNI KTPVKIAIFT LICTQLMNLA  
 401 FIGPLKHVGL SLAIGLGACI NAGLLFYLLR RHGIYQPGKG WAAFLAKMLL  
 451 SLAVMCGGLW AAQAYLPFEW AHAGGMRKAG QLCILIAVGG GLYFASLAL  
 501 GFRPRHFRRV EN\*

10 Computer analysis of this amino acid sequence gave the following results:

Homology with the MviN virulence factor of *S. typhimurium* (accession number P37169)

ORF20 and MviN proteins show 63% aa identity in 440aa overlap:

Orf20 1 MNMLGALAKVGSMTMVSRLGVFVRDVIARAFAFGAGMATDAFFVAFKLPNLLRRVFAEGAF 60  
 MN+L +LA V S+TM SRVLGF RD ++AR FGAGMATDAFFVAFKLPNLLRR+FAEGAF  
 MviN 14 MNLLKSLAAVSSMTMFSRLGVFARDIAVARIFGAGMATDAFFVAFKLPNLLRRIFAEGAF 73  
 Orf20 61 AQAFVPILAIEYKETSKEAXEAFIRHVAGMLSFLVIVTALGILAAPWVIYVSAPSFAQD 120  
 +QAFVPILAIEYK + +EA F+ +V+G+L+ L +VT G+LAAPWVI V+AP FA  
 MviN 74 SQAFVPILAIEYKSKQGEATRIFVAYVSGLLTLALAVVTVAGMLAAPWVIMVTAPGFADT 133  
 Orf20 121 ADKFQLSIDLLRITFPYILLISLSSEFVGSVLNSYHKFGIPAFTPXFLNVSFIVFALFFVP 180  
 ADKF L+ LLRITFPYILLISL+S VG++LN++++F IPAF P FLN+S I FALF P  
 MviN 134 ADKFALTTLRITFPYILLISLASLVGAILNTWNRFSIPAFAPTFLNISMIGFALFAAP 193  
 Orf20 181 YFDPPTAXAWAVFVGGILQLXFLPWLAKLGFLKLPKLSFKDAAVNRVMKQMAPAILGV 240  
 YF+PPV A AWAV VGG+LQL +QLP+L K+G L LP+++F+D RV+KQM PAILGV  
 MviN 194 YFNPPVLALAWAVTVGGVLQLVYQLPYLKKIGMLVLPRIINFRDTGAMRVVKQMGPAILGV 253  
 Orf20 241 SVAQVSLVINTIFASYLQSGSVSWMYADRMELPSGVLGAAALGTILLPTLSKHSANQDT 300  
 SV+Q+SL+INTIFAS+L SGSVSWMYADR+ME PSGVLG ALGTILLP+LSK A+ +  
 MviN 254 SVSISLSLIINTIFASFLASGSVSWMYADRLMEFSPGVLGVALGTILLPSLSKSFASGNH 313  
 Orf20 301 EQFSALLDWGLRLCMLLTLPAAVGLAVLSFPLVATLFMYRXTLFDAQMTQHALIAYSFG 360  
 +++ L+DWGLRLC LL LP+AV L +L+ PL +LF Y FT FDA MTQ ALIAYS G  
 MviN 314 DEYCRMLDWGLRLCFLALPSAVALGILAKPLTVSLFQYGKFTAFDAAMTQRALIAYSVG 373  
 Orf20 361 LIGLIMIKVLAPGFYARQNIIXPVKIAIFTLIXQLMNLXFXXXXXXXXXXXXXXXXXXCI 420  
 LIGLI++KVLAPGFY+RQ+I PVKIAI TLI QLMNL F C+  
 MviN 374 LIGLIVVKVLAPGFYSRQDIKTPVKIAIVTLIMTQLMNLAFIGPLKHAGLSLSIGLAACL 433  
 Orf20 421 NAGLLFYLLRRHGIYQXPQG 440  
 NA LL++ LR+ I+ P G  
 MviN 434 NASLLYWQLRKQNIPTQPG 453

45 Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF20 shows 93.5% identity over a 447aa overlap with an ORF (ORF20a) from strain A of *N. meningitidis*:

orf20.pep MNMLGALAKVGSMTMVSRLGVFVRDVIARAFAFGAGMATDAFFVAFKLPNLLRRVFAEGAF 60  
 orf20a MNMLGALVKVGSMTMVSRLGVFVRDVIARAFAFGAGMATDAFFVAFKLPNLLRRVFAEGAF 60  
 orf20.pep AQAFVPILAIEYKETSKEAXEAFIRHVAGMLSFLVIVTALGILAAPWVIYVSAPSFAQD 120  
 orf20a AQAFVPILAIEYKETSKEATEAFIRHVAGMLSFLVIVTALGILAAPWVIYVSAPGFAKD 120

		130	140	150	160	170	180
	orf20.pep	ADKFQLSIDLLRITFPYILLISLSSFVGSVLNSYHKFGIPAFTPXFLNVSFIVFALFFVP					
	orf20a	ADKFQLSIDLLRITFPYILLISLSSFVGSVLNSYHKFSIPAFTPFTFLNVSFIVFALFFVP					
5		130	140	150	160	170	180
	orf20.pep	YFDPPTAXAWAVFVGGILQLXFQLPWLAKLGLKLPKLSFKDAAVNRVMKQMAPAILGV					
10	orf20a	YFDPPTAXAWAVFVGGILQLGFQLPWLAKLGLKLPKLSFKDAAVNRVMKQMAPAILGV					
		190	200	210	220	230	240
	orf20.pep	SVAQVSLVINTIFASYLQSGSVSWMYADRMELPSGVLGAALGTILLPTLSKHSANQDT					
15	orf20a	SVAQISLVINTIFASYLQSGSVSWMYADRMELPGGVLGAALGTILLPTLSKHSANQDT					
		250	260	270	280	290	300
	orf20.pep	EQFSALLDWGLRLCMLLTLPAAVGLAVLSFPLVATLFMYRXFTLFDAQMTQHALIAYSFG					
20	orf20a	EQFSALLDWGLRLCMLLTLPAAVGMVAVLSFPLVATLFMYREFTLFDAQMTQHALIAYSFG					
		310	320	330	340	350	360
	orf20.pep	LIGLIMIKVLAPGFYARQNIKXPVKIAIFTLICXQLMNLXFXGPLXXIGLSLAIGLGACI					
25	orf20a	LIGLIMIKVLAPGFYARQNIKTPVKIAIFTLICTQLMNLAFIGPLKHVGLSLAIGLGACI					
		370	380	390	400	410	420
	orf20.pep	NAGLLFYLLRRHGIYQPGXQLGSLVXQKCCSRSPX					
30	orf20a	NAGLLFYLLRRHGIYQPGKWAFLAKMLLSLAVMGGGLYAAQIWLFPDWAHAGGMQKAA					
		430	440	450	460	470	480

The complete length ORF20a nucleotide sequence <SEQ ID 117> is:

	1	ATGAATATGC	TGGGAGCTTT	GGTAAAAGTC	GGCAGCCTGA	CGATGGTGTC
	51	GCGCGTTTTG	GGATTGTGTC	GCGATACGGT	CATTGCGCGC	GCATTCGGCG
	101	CAGGCATGGC	GACGGATGCG	TTCTTTGTCTG	CGTTCAAACT	GCCCAACCTG
40	151	CTTCGCCGCG	TGTTTGCGGA	GGGGCGCTTT	GCCCAAGCGT	TTGTGCCGAT
	201	TTTGCGCGAA	TATAAGGAAA	CGCGTTCTAA	AGAGGCGACG	GAGGCTTTTA
	251	TCCGCCATGT	GGCGGGGATG	CTGTCGTTTG	TACTGGTCAT	CGTTACCGCG
	301	CTGGGCATAC	TTGCCGCGCC	TTGGGTGATT	TATGTTTCCG	CACCGGTTT
	351	TGCCAAAGAT	GCCGACAAAT	TTCAGCTCTC	TATCGATTG	CTGCGGATTA
45	401	CGTTTCCTTA	TATCTTATTG	ATTTCACCTT	CCTCTTTTGT	CGGCTCGGTA
	451	CTCAATTCTT	ATCATAAATT	CAGCATTCCT	CGGTTTACGC	CCACGTTCTT
	501	GAACGTGTCT	TTTATCGTAT	TCGCGCTGTT	TTTCGTGCCG	TATTTTCGATC
	551	CTCCCGTTAC	CGCGCTGGCT	TGGGCGGTTT	TTGTCCGCGG	CATTTTGCAA
	601	CTCGGCTTCC	AACTGCCCTG	GCTGGCGAAA	CTGGGTTTTT	TGAAACTGCC
50	651	CAAAGTGAGT	TTCAAAGATG	CGGCGGTCAA	CCGCGTGATG	AAACAGATGG
	701	CGCCTGCGAT	TTTGGGCGTG	AGCGTGGCGC	AGATTTCTTT	GGTGATCAAC
	751	ACGATTTTCG	CGTCTTATCT	GCAATCGGGC	AGCGTTTCAT	GGATGTATTA
	801	CGCCGACCGC	ATGATGGAAC	TGCCCGGCGG	CGTGCTGGGG	GCGGCACTCG
	851	GTACGATTTT	GCTGCCGACT	TTGTCCAAAC	ACTCGGCAAA	CCAAGATACG
55	901	GAACAGTTTT	CCGCCCTGCT	CGACTGGGGT	TTGCGCNTGT	GCATGCTGCT
	951	GACGCTGCCG	GCGGCGGTCT	GAATGGCGGT	GTTGTCTGTT	CCGCTGGTGG
	1001	CAACCTTGTT	TATGTACCGA	GAATTCACGC	TGTTTGACGC	GCAGATGACG
	1051	CAACACGCGC	TGATTGCCTA	TTCTTTCCGT	TTAATCGGTT	TAATCATGAT
	1101	TAAAGTGTTG	GCGCCCGGCT	TTTATGCGCG	GCAAAACATC	AAAACGCCCG
60	1151	TCAAATCGC	CATCTTCACG	CTCATTTGCA	CGCAGTTGAT	GAACCTTGCC
	1201	TTTATCGGCC	CACGTAAACA	CGTCGGACTT	TGCTTGCCA	TGGTCTGGG
	1251	CGCGTGTATC	AATGCCGGAT	TGTTGTTTGA	CCTGTTGCGC	AGACACGGTA
	1301	TTTACCAACC	TGGCAAGGGT	TGGGACGCGT	TCTTGGCAAA	AATGCTGCTC
	1351	TCGCTCGCCG	TGATGGGAGG	CGGCCTGTAT	GCCGCCCAAA	TCTGGCTGCC
65	1401	GTTCGACTGG	GCACACGCCG	GCGGAATGCA	AAAGGCCGCC	CGGCTCTTCA
	1451	TCCTGATTGC	CGTCGGCGGC	GGACTGTATT	TCGCATCACT	GGCGGCTTTG
	1501	GGCTTCCGTC	CGCGCAATTT	CAAACGCGTG	GAAAGCTGA	



This encodes a protein having amino acid sequence <SEQ ID 118>:

1 MNMLGALVKV GSLTMVSRVL GFVRDVIAR AFGAGMATDA FFAFKLPNL  
51 LRRVFAEGAF AQAFVPILAE YKETRSKEAT EAFIRHVAGM LSFVLVIVTA  
101 LGILAAPWVI YVSAPGFAKD ADKFQLSIDL LRITFPYILL ISLSSFVGSV  
5 LNSYHKFSIP AFTPTFLNVS FIVFALFFVP YFDPPTALA WAVEVGGILQ  
201 LGFQLPWLAK LGFLKLPKLS FKDAAVNRVM KQMAPAILGV SVAQISLVIN  
251 TIFASYLQSG SVSWMYADR MMELPGGVLG AALGTILLPT LSKHSANQDT  
301 EQFSALLDWG LRXCMLLTLPA AVGMAVLSF PLVATLFMYR EFTLFDAQMT  
351 QHALIAYSFG LIGLIMIKVL APGFYARQNI KTPVKIAIFT LICTQLMNL  
10 FIGPLKHVGL SLAIGLGACI NAGLLFYLLR RHGIYQPGKG WAAFLAKMLL  
451 SLAVMGGGLY AAQIWLPPDW AHAGGMQKAA RLFILIAVGG GLYFASLAAL  
501 GFRPRHFRRV ES\*

ORF20a and ORF20-1 show 96.5% identity in 512 aa overlap:

15 orf20a.pep MNMLGALVKVGSGLTMVSRVLGFVRDVIARAFGAGMATDAFFVAFKLPNLLRRVFAEGAF  
orf20-1 MNMLGALAKVGSGLTMVSRVLGFVRDVIARAFGAGMATDAFFVAFKLPNLLRRVFAEGAF  
20 orf20a.pep AQAFVPILAEYKETRSKEATEAFIRHVAGMLSFVLVIVTALGILAAPWVIYVSAPGFAKD  
orf20-1 AQAFVPILAEYKETRSKEAAEAFIRHVAGMLSFVLVIVTALGILAAPWVIYVSAPGFAQD  
25 orf20a.pep ADKFQLSIDLLRITFPYILLISLSSFVGSVLNSYHKFSIPAFPTPTFLNVSFIVFALFFVP  
orf20-1 ADKFQLSIDLLRITFPYILLISLSSFVGSVLNSYHKFGIPAFPTPTFLNVSFIVFALFFVP  
30 orf20a.pep YFDPPTALAWAVFVGGILQLGFQLPWLAKLGFLKLPKLSFKDAAVNRVMKQMAPAILGV  
orf20-1 YFDPPTALAWAVFVGGILQLGFQLPWLAKLGFLKLPKLSFKDAAVNRVMKQMAPAILGV  
35 orf20a.pep SVAQISLVINTIFASYLQSGSVSWMYADRMELPGGVLGAALGTILLPTLSKHSANQDT  
orf20-1 SVAQVSLVINTIFASYLQSGSVSWMYADRMELPSGVLGAALGTILLPTLSKHSANQDT  
40 orf20a.pep EQFSALLDWGLRXCMLLTLPAAVGMAVLSFPLVATLFMYREFTLFDAQMTQHALIAYSFG  
orf20-1 EQFSALLDWGLRLCMLLTLPAAVGLAVLSFPLVATLFMYREFTLFDAQMTQHALIAYSFG  
45 orf20a.pep LIGLIMIKVLAPGFYARQNIKTPVKIAIFTLICTQLMNLAFIGPLKHVGLSLAIGLGACI  
orf20-1 LIGLIMIKVLAPGFYARQNIKTPVKIAIFTLICTQLMNLAFIGPLKHVGLSLAIGLGACI  
50 orf20a.pep NAGLLFYLLRRHGIYQPGKGWAAFLAKMLLSLAVMGGGLYAAQIWLPPDWAHAGGMQKAA  
orf20-1 NAGLLFYLLRRHGIYQPGKGWAAFLAKMLLSLAVMCGGLWAAQAYLPFEWAHAGGMRKAG  
55 orf20a.pep RLFILIAVGGGLYFASLAALGFRPRHFKRRESX  
orf20-1 QLCILIAVGGGLYFASLAALGFRPRHFKRVENX  
60 orf20a.pep  
orf20-1

Homology with a predicted ORF from *N.gonorrhoeae*

ORF20 shows 92.1% identity over a 454aa overlap with a predicted ORF (ORF20ng) from *N.gonorrhoeae*:

5	orf20.pep	MNMLGALAKVGS LTMVSRVLGFVRD TVIARAFGAGMATDAFFVAFKLPNLLRRVFAEGAF	60
	orf20ng	MNMLGALAKVGS LTMVSRVLGFVRD TVIARAFGAGMATDAFFVAFKLPNLLRRVFAEGAF	60
10	orf20.pep	AQAFVPILAEYKETRSKEAXEAFIRHVAGMLS FVLVIVTALGILAAPWVIYVSAPSFAQD	120
	orf20ng	AQAFVPILAEYKETRSKEATEAFIRHVAGMLS FVLIVVTALGILAAPWVIYVSAPGFTKD	120
15	orf20.pep	ADKFQLSIDLRLITFPYILLISLSSFVGSVLNSYHKFGIPAFTPXFLNVSFIVFALFFVP	180
	orf20ng	ADKFQLSISLLRITFPYILLISLSSFVGSILNSYHKFGIPAFTPFLNISFIVFALFFVP	180
20	orf20.pep	YFDPVPTAXAWAVFVG GILQLXFLPWLAKLGF LKLPKLSFKDAAVNRVMKQMAPAILGV	240
	orf20ng	YFDPVPTALAWAVFVG GILQLGF LKLPWLAKLGF LKLPKLNFKDAAVNRVMKQMAPAILGV	240
25	orf20.pep	SVAQVSLVINTIFASYLQSGSVSWMYADRMELPSGVLGAALGTILLPTLSKHSANQDT	300
	orf20ng	SVAQISLVINTIFASYLQSGSVSWMYADRMELPGGVLGAALGTILLPTLSKHSANQDT	300
30	orf20.pep	EQFSALLDWGLRLCMLLTLPAAVGLAVLSFPLVATLFMYRXFTLFDAQMTQHAIAYSFG	360
	orf20ng	EQFSALLDWGLRLCMLLTLPAAAGLAVLSFPLVATLFMYREFTLFDAQMTQHAIAYSFG	360
35	orf20.pep	LIGLIMIKVLAPGFYARQNI XPKIAIFTLICKQLMNLXFXGPLXXIGLSLAIGLGACI	420
	orf20ng	LIGLIMIKVLASGFYARQNI KTPVKIAIFTLICTQLMNLAFIGPLKHAGLSLAIGLGACI	420
	orf20.pep	NAGLLFYLLRRHGIYQXPXQGLGSLVXQKCCSRSP	454
	orf20ng	NAGLLFFLFRKHGIYRPGQGLGQPSWRKCCSRSP	454

An ORF20ng nucleotide sequence <SEQ ID 119> was predicted to encode a protein having amino acid sequence <SEQ ID 120>:

1	MNMLGALAKV	GSLTMVSRVL	GFVRD TVIAR	AFGAGMATDA	FFVAFKLPNL
51	LRVFAEGAF	AQAFVPILAE	YKETRSKEAT	EAFIRHVAGM	LSFVLIVVTA
101	LGILAAPWVI	YVSAPGFTKD	ADKFQLSISL	LRITFPYILL	ISLSSFVGS
151	LNSYHKFGIP	AFTPTFLNIS	FIVFALFFVP	YFDPVPTALA	WAVFVG GILQ
201	LGFLPWLAK	LGFLKLPKLN	FKDAAVNRVM	KQMAPAILGV	SVAQISLVIN
251	TIFASYLQSG	SVSWMYADR	MMELPGGVLG	AALGTILLPT	LSKHSANQDT
301	EQFSALLDWG	LRLCMLLTLP	AAAGLAVLSF	PLVATLFMYR	EFTLFDAQMT
351	QHAIAYSFG	LIGLIMIKVL	ASGFYARQNI	KTPVKIAIFT	LICTQLMNL
401	FIGPLKHAGL	SLAIGLGACI	NAGLLFFLFR	KHGIYRPGQG	LGQPSWRKCC
451	SRSP*				

Further DNA sequence analysis revealed the following DNA sequence <SEQ ID 121>:

1	ATGAATATGC	TTGGAGCTTT	GGCAAAAGTC	GGCAGCCTGA	CGATGGTGTC
51	GCGCGTTTTG	GGATTGTGTC	GCGATACGGT	CATTGCGCGG	GCATTCGGCG
101	CGGGTATGGC	GACGGATGCG	TTTTTTGTCTG	CGTTCAAAC	GCCCAACCTG
151	CTTCGCCGCG	TGTTTTCGGA	GGGGCGGTTT	GCCCAAGCGT	TTGTGCCGAT
201	TTTGCGGAA	TATAAGGAAA	CGCGTTCTAA	AGAGGCGAcg	gAGGCTTTTA
251	TCCGCCACGt	tgccgggAatg	CTGTCGTTTG	TGCTGATcgt	cGttacCGCG
301	CTGGGCATAC	TTGCCGCGcc	tTGGGTGATT	TATGTTtccg	CgcccGGCTT
351	TACCAAAGAC	GCGGACAAGT	TCCAACCTTC	CATCAGCCTG	CTGCGGATTA
401	CGTTTCCTTA	TATATTATTG	ATTTCTTTGT	CTTCTTTTGT	CGGCTCGATA
451	CTCAATTCCt	ACCATAAGTT	CGGCATTCCC	CGGTTTACG	CCACGTTTTT
501	AAACATCTCT	TTTATCGTAT	TCGCACTGTT	TTTCGTGCCG	TATTTTCGATC
551	CGCCCGTTAC	CGCGCTGGCG	TGGCGGTTT	TTGTCGCGCG	TATTTTGCAG

5  
10  
15  
20

```

601 CTCGGTTTCC AACTGCCGTG GCTGGCGAAA CTGGGCTTTT TGAAACTGCC
651 CAAACTGAAT TTCAAAGATG CGGCGGTCAA CCGCGTCATG AAACAGATGG
701 CGCCTGCGAT TTTGGGCGTG agcgtGGCGC AAATTTCTTT GgttATCAAC
751 ACGATTTTCG CGTCTTATCT GCAATCGGGC AGCGTTTCAT GGATGTatta
801 cgCCGACCGC ATGATGGAGc tgcgccGGGG CGTGCTGGGG GCTGCACTCG
851 GTACAAATTTT GCTGCCGACT TTGTCCAAAC ACTCGGCAAA CCAAGATACG
901 GAACAGTTTTT CCGCCCTGCT CGACTGGGGT TTGCGCCTGT GCATGCTGCT
951 GACGCTGCCG GCGGCGGccg GACTGGCGGT ATTGTGTTT CCGCTGGTGG
1001 CGACGCTGTT TATGTACCGA GAATTCACGC TGTTCGACGC ACAAATGACG
1051 CAACACGCGC TGATTGCCTA TTCTTTTCGGT TTAATCGGTT TAATTATGAT
1101 TAAAGTGTG GCATCCGGCT TTTATGCGCG GCAAAACATC AAAACGCCCG
1151 TCAAAATCGC CATCTTCACG CTCATCTGCA CGCAGTTGAT GAACCTCGCC
1201 TTTATCGGTC CGTTGAAACA CGCCGGGCTT TCGCTCGCCA TCGGCCTGGG
1251 CGCGTGCATC AACGCCGGAT TGTGTCTCTT CCTGTTGCGC AAACACGGTA
1301 TTTACCGGCC cggcaggggt tgggcggcgt TCTTGGCGAA AATGCTGCTC
1351 GCGCTCGCCG TGATGTGCGG CGGACTGTGG GCGGCGCAGG CTTGCCTGCC
1401 GTTCGAATGG GCGCAGCGCG CGGGAATGCG GAAAGCGGGG CAGCTCTGCA
1451 TCCTGATTGC CGTCGGCGGC GGACTGTATT TCGCATCTCT GCGGCTTTG
1501 GGCTTCCGTC CCGGCCATTT CAAACGCGTG GAAAGCTGA

```

20 This encodes the following amino acid sequence <SEQ ID 122; ORF20ng-1>:

1 MNMLGALAKV GSMTVSRVL GFVRDVIAR AFGAGMATDA FFAVFKLPNL  
51 LRRVFAEGAF AQAFVPILAE YKETRSKEAT EAFIRHVAGM LSFVLIVVTA  
101 LGILAAPWVI YVSAPGFTKD ADKFQLSISL LRITFPYILL ISLSSFVSGI  
151 LNSYHKFGIP AFTPTFLNIS FIVEALFFVP YFDPPTVTLA WAVFVGILQ  
201 LGFQLPWLAK LGFLKLPKLN FKDAAVNRVM KQMAPAILGV SVAQISLVIN  
251 TIFASYLQSG SVSWMYADR MMELRRGVLG AALGTILLPT LSKHSANQDT  
301 EQFSALLDWG LRLCMLLTLF AAAGLAVLSF PLVATLFMYR EFTLFDAQMT  
351 QHALIAYSFG LIGLIMIKVL ASGFYARQNI KTPVKIAIFT LICTQLMNL  
401 FIGPLKHAGL SLAIGLGACI NAGLLFFLLR KHGIYRPGRG WAAFLAKMLL  
451 ALAVMCGGLW AAQACLPEW AHAGGMRKAG QLCILIAVGG GLYFASLAAL  
501 GFRPRHFKRV ES\*

ORF20ng-1 and ORF20-1 show 95.7% identity in 512 aa overlap:

35  
40  
45  
50  
55  
60  
65

```

      10      20      30      40      50      60
orf20-1.pep MNMLGALAKVGSMTVSRVLGFVRDVIARAFAFGAGMATDAFFVAFKLPNLLRRVFAEGAF
      |||
orf20ng-1 MNMLGALAKVGSMTVSRVLGFVRDVIARAFAFGAGMATDAFFVAFKLPNLLRRVFAEGAF
      10      20      30      40      50      60

      70      80      90      100     110     120
orf20-1.pep AQAFVPILAEYKETRSKEAAEAFIRHVAGMLSFVLIVVTALGILAAPWVIYVSAPGFAQD
      |||
orf20ng-1 AQAFVPILAEYKETRSKEATEAFIRHVAGMLSFVLIVVTALGILAAPWVIYVSAPGFTKD
      70      80      90      100     110     120

      130     140     150     160     170     180
orf20-1.pep ADKFQLSIDLLRITFPYILLISLSSFVGSVLNSYHKFGIPAFPTFLNVSFIVFALFFVP
      |||
orf20ng-1 ADKFQLSISLLRITFPYILLISLSSFVGSILNSYHKFGIPAFPTFLNISFIVFALFFVP
      130     140     150     160     170     180

      190     200     210     220     230     240
orf20-1.pep YFDPPTVTLAWAVFVGILQLGFQLPWLAKLGLKLPKLSFKDAAVNRVMKQMAPAILGV
      |||
orf20ng-1 YFDPPTVTLAWAVFVGILQLGFQLPWLAKLGLKLPKLNFKDAAVNRVMKQMAPAILGV
      190     200     210     220     230     240

      250     260     270     280     290     300
orf20-1.pep SVAQVSLVINTIFASYLQSGSVSWMYADRMELPSGVLGAALGTILLPTLSKHSANQDT
      |||
orf20ng-1 SVAQISLVINTIFASYLQSGSVSWMYADRMELRRGVLGAALGTILLPTLSKHSANQDT
      250     260     270     280     290     300

      310     320     330     340     350     360
orf20-1.pep EQFSALLDWGLRLCMLLTLPAAGLAVLSFPLVATLFMYREFTLFDAQMTQHALIAYSFG
      |||
orf20ng-1 EQFSALLDWGLRLCMLLTLPAAGLAVLSFPLVATLFMYREFTLFDAQMTQHALIAYSFG

```

		310	320	330	340	350	360
		370	380	390	400	410	420
5	orf20-1.pep	LIGLIMIKVLAPGFYARQNIKTPVKIAIFTLICTQLMNLAFIGPLKHVGLSLAIGLGACI					
	orf20ng-1	LIGLIMIKVLASGFYARQNIKTPVKIAIFTLICTQLMNLAFIGPLKHAGLSLAIGLGACI					
		370	380	390	400	410	420
10	orf20-1.pep	NAGLLFYLLRRHGIYQPGKGWAAFLAKMLLSLAVMCGGLWAAQAYLPFEWAHAGGMRKAG					
	orf20ng-1	NAGLLFFLLRKHGIYRPGRGWAAFLAKMLLALAVMCGGLWAAQACLPFEWAHAGGMRKAG					
		430	440	450	460	470	480
15	orf20-1.pep	QLCILIAVGGGLYFASLAALGFRPRHFKRVENX					
	orf20ng-1	QLCILIAVGGGLYFASLAALGFRPRHFKRVESX					
		490	500	510			

20 In addition, ORF20ng-1 shows significant homology with a virulence factor of *S.typhimurium*:

sp|P37169|MVIN\_SALTY VIRULENCE FACTOR MVIN pir||S40271 mviN protein - Salmonella  
 typhimurium gi|438252 (Z26133) mviB gene product [Salmonella typhimurium]  
 gnl|PID|d1005521 (D25292) ORF2 [Salmonella typhimurium] Length = 524  
 Score = 1573 (750.1 bits), Expect = 1.1e-220, Sum P(2) = 1.1e-220  
 Identities = 309/467 (66%), Positives = 368/467 (78%)

25 Query: 1 MNMLGALAKVGS LTMVSRVLGFVRD TVIARA FGAGMATDAFFVAFKLPNLLRRVFAEGAF 60  
 MN+L +LA V S+TM SRVLGF RD ++AR FGAGMATDAFFVAFKLPNLLRR+FAEGAF  
 30 Sbjct: 14 MNLLKSLAAVSSMTMF SRVLGFARDAIVARIFGAGMATDAFFVAFKLPNLLRRIFAEGAF 73

35 Query: 61 AQAFVPILA EYKETSKEATEAFIRHVAGMLS FVLIVVTALGILAAPWVIYVSAPGFTKD 120  
 +QAFVPILA EYK + +EAT F+ +V+G+L+ L VVT G+LAAPWVI V+APGF  
 Sbjct: 74 SQAFVPILA EYKSKQGE EATRIFVAYVSGLLTLALAVVTVAGMLAAPWVIMVAPGFADT 133

40 Query: 121 ADKFQLSISLLRITFPYILLISLS SFGVGSILNSYHKFGIPAF TPTFLNISFIVFALFFVP 180  
 ADKF L+ LLRITFPYILLISL+S VG+ILN++++F IPAF PTFNLIS I FALF P  
 Sbjct: 134 ADKFALTTQLLRITFPYILLISLASLVGAILNTWNRFSIPAFAPTFLNISMIGFALFAAP 193

45 Query: 181 YFDP PVTALAWAVFVG GILQLGFLPWLAKLGFLKLPKLNFKDAAVNRVMKQMAPAILGV 240  
 YF+PPV ALAWAV VGG+LQL +QLP+L K+G L LP++NF+D RV+KQM PAILGV  
 Sbjct: 194 YFNPPVLALAWAVTVGGV LQVLYQLPYLKKIGMLVLP RINFRDTGAMRVVKQMGPAAILGV 253

50 Query: 241 SVAQISLVINTIFASYLQSGSVSWMY YADRMELRRGV LGAALGTILLPTLSKHSANQDT 300  
 SV+QISL+INTIFAS+L SGVSWSMY YADR+ME GVLG ALGTILLP+LSK A+ +  
 Sbjct: 254 SVSQISLIINTIFASFLASGVSWSMY YADRLMEFPGSVLGVALGTILLPSLSKSFASGNH 313

55 Query: 301 EQFSALLDWGLRLCMLTLPAAGLAVLSFPLVATLFMYREFTLFDAQMTQHAIAYSFG 360  
 +++ L+DWGLRLC LL LP+A L +L+ PL +LF Y +FT FDA MTQ ALIAYS G  
 Sbjct: 314 DEYCRLMDWGLRLCFLALPSAVALGILAKPLTVSLFQYKFTAFDAAMTQRALIAYSVG 373

60 Query: 361 LIGLIMIKVLASGFYARQNIKTPVKIAIFTLICTQLMNLAFIGPLKHAGLSLAIGLGACI 420  
 LIGLI++KVL A GFY+RQ+IKTPVKIAI TLI TQLMNLAFIGPLKHAGLSL+IGL AC+  
 Sbjct: 374 LIGLIVVKVLAPGFYSRQDIKTPVKIAIVTLIMTQLMNLAFIGPLKHAGLSLSIGLAAC 433

65 Query: 421 NAGLLFFLLRKHGIYRPGRGWXXXXXXXXXXXXVMCGGLWAAQACLP 467  
 NA LL++ LRK I+ P GW VM L+ +P  
 Sbjct: 434 NASLLYWQLRKQNI FT PQPGWMMWFLMRIISVLVMAAVLFGVLHIMP 480

Score = 70 (33.4 bits), Expect = 1.1e-220, Sum P(2) = 1.1e-220  
 Identities = 14/41 (34%), Positives = 23/41 (56%)

Query: 469 EWAHAGGMRKAGQLCILIAVGGGLYFASLAALGFRPRHFKR 509  
 EW+ + + +L ++ G YFA+LA LGF+ + F R  
 Sbjct: 481 EWSQGSMLWRLLRLMAVVIAGIAAYFAALAVLGFVKVKEFVR 521

Based on this analysis, including the homology with a virulence factor from *S.typhimurium*, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 15

5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 123>:

```

1   atGATTAAAA TCAAAAAAGG TCTAAACCTG CCCATCGCGG GCAGACCGGA
51  GCAAGCCGTT TACGACGGCC CGGCCaTTAC CGAAGtCGCG TTGCTTGGCG
101 AAGAATATGC CGGTATGCGC CCCTCGATGA AAGTCAAGGA AGGCGATGCC
151 GTcAAAAAAG GCCAAGTGCT GTTTGAAGAC AAAAAGAATC CGGGCGTGGT
10  201 GTTTACTGCG CCGGCTTCAG GcAAAATCGC CGCGATTAC CGTGGCGAAA
251 AGCGCGTACT TCAGTCAGTC GTGATTGCCG TTGAAGGCAA CGACGAAATC
301 GAGTTTGAAC GCTACGCACC TGAAGCGCTG GCAAACCTAA GCGGCGAAGA
351 AGTGCGCCGC AACCTGATCC AATCCGGTTT GTGGACTGCG CTGCGCACCC
401 GTCCGTTTCA CAAAATTCCT GCCGTCGATG CCGAGCCGTT CGCCATCTTC
15  451 GTCAATGCGA tGGACACCAA TCCG..

```

This corresponds to the amino acid sequence <SEQ ID 124; ORF22>:

```

1   MIKIKKGLNL PIAGRPEQAV YDGPATEVA LLGEEYAGMR PSMKVKEGDA
51  VKKGQVLFED KKNPGVVFTA PASGKIAAIH RGEKRVLQSV VIAVEYNDEI
20  101 EFERYAPEAL ANLSGEEVRR NLIQSGLWTA LRTRPFSKIP AVDAEPFAIF
151 VNAMDTPNP..

```

Further work revealed the complete nucleotide sequence <SEQ ID 125>:

```

1   ATGATTAAAA TCAAAAAAGG TCTAAACCTG CCCATCGCGG GCAGACCGGA
51  GCAAGCCGTT TACGACGGCC CGGCCATTAC CGAAGTTCGCG TTGCTTGGCG
25  101 AAGAATATGC CGGTATGCGC CCCTCGATGA AAGTCAAGGA AGGCGATGCC
151 GTCAAAAAAG GCCAAGTGCT GTTTGAAGAC AAAAAGAATC CGGGCGTGGT
201 GTTTACTGCG CCGGCTTCAG GCAAAATCGC CGCGATTAC CGTGGCGAAA
251 AGCGCGTACT TCAGTCAGTC GTGATTGCCG TTGAAGGCAA CGACGAAATC
301 GAGTTTGAAC GCTACGCACC TGAAGCGCTG GCAAACCTAA GCGGCGAAGA
351 AGTGCGCCGC AACCTGATCC AATCCGGTTT GTGGACTGCG CTGCGCACCC
401 GTCCGTTTCA CAAAATTCCT GCCGTCGATG CCGAGCCGTT CGCCATCTTC
451 GTCAATGCGA TGGACACCAA TCCGCTGGCT GCCGACCTA CGGTCATTAT
501 CAAAGAAGCC GCCGAGGATT TCAAACGCGG CCTGTTGGTA TTGAGCCGTT
551 TGACCGAAGC CAAAATCCAT GTTTGTAAGG CAGCTGGCGC AGACGTGCCG
601 TCTGAAAATG CTGCCAACAT CGAAACACAT GAATTTCGGC GCCCGCATCC
35  651 TGCCGGTTTG AGTGGCACGC ACATTCAATT CATCGAGCCG GTCGGCGCGA
701 ATAAACCCTG GTGGACCATC AATTATCAAG ATGTAATTAC CATTGGCCGT
751 TTGTTTGCAA CAGGCCGTCT GAACACCGAG CGCGTGATTG CCCTAGGTGG
801 TTCTCAAGTC AACAAACCGC GCCTCTTGCG TACCGTTTTG GGTGCGAAAG
40  851 TATCGCAAAT TACTGCGGGC GAATTGGTTG ACACAGACAA CCGCGTGATT
901 TCCGGTTTCG TATTGAACGG CGCGATTACA CAAGGCGCGC ACGATTATTT
951 GGGACGCTAC CACAATCAGA TTTCGTTAT CGAAGAAGGC CGCAGCAAAG
1001 AGCTGTTTCG CTGGGTTGCG CCGCAGCCGG ACAATACTC CATCAGCGT
1051 ACAACCCTCG GCCATTTCTT GAAAAACAAA CTCTCAAGT TCAACACAGC
45  1101 CGTCAACGGC GCGCAGCCGC CCATGGTGCC GATTGGTACT TACGAGCGCG
1151 TGATGCCCTT GGATATCTG CCCACCCTGC TTTTGCGCGA TTTAATCGTC
1201 GCGGATACCG ACAGCGCGCA GGCATTGGGT TGCTTGGAAT TGGACGAAGA
1251 AGACCTCGCT TTGTGACGCT TCGTCTGCCG GGGCAAATAC GAATACGGCC
1301 CGCTGTTGCG CAAAGTGCTG GAAACCATTG AGAAGGAAGG CTGA

```

This corresponds to the amino acid sequence <SEQ ID 126; ORF22-1>:

```

50  1   MIKIKKGLNL PIAGRPEQAV YDGPATEVA LLGEEYAGMR PSMKVKEGDA
51  VKKGQVLFED KKNPGVVFTA PASGKIAAIH RGEKRVLQSV VIAVEYNDEI
101 EFERYAPEAL ANLSGEEVRR NLIQSGLWTA LRTRPFSKIP AVDAEPFAIF
151 VNAMDTPNPLA ADPTVIIKEA AEDFKRGLLV LSLRTERKIH VCKAAGADV
55  201 SENAANIETH EFGGPHFAGL SGTHIHFIET VGANKTVWTI NYQDVITIGR
251 LFATGRLNTE RVIALGGSQV NKPRLLRTVL GAKVSQITAG ELVDTDNRVI

```

301 SGSVLNGAIT QGAHDYLGRY HNQISVIEEG RSKELFGWVA POPDKYSITR  
 351 TTLGHFLKNK LFKENTAVNG GDRAMVPIGT YERVMPLDIL PTLLLRDLIV  
 401 GDTDSAQALG CLELDEEDLA LCSFVCPGKY EYGPLLRLKVL ETIEKEG\*

Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 127>:

5 1 ATGATTAAAA TCAAAAAAGG TCTAACCTG CCCATCGCGG GCAGACCGGA  
 51 GCAAGTCATT TATGACGGGC CCGTCATTAC CGAAGTCGCG TTGCTTGGCG  
 101 AAGAATATGC CCGTATGCGC CCCTNGATGA AAGTCAAGGA AGGCGATGCC  
 151 GTCAAAAAAG GCCAAGTGCT GTTTGAAGAC AAAAAGNATC CGGGCGTGGT  
 201 GTTTACCGCG CCNGTTTCAG GCAAAATCGC CGCCATCCAT CGCGGCGAAA  
 10 251 AGCGCGTACT TCAGTCGGTC GTGATTGCCG TTGAAGGCAA CGACGAAATC  
 301 GAGTTCGAAC GCTACGCGCC CGAAGCGTTG GCAAACCTAA GCGGCGANGA  
 351 ANTNNGNNGC AATCTGATCC AATCCGGTTT GTGGACTGCG CTGCGTANCC  
 401 TCCGTTTCAG CAAAATCCCT GCCGTCGATG CCGAGCCGTT CGCCATCTTC  
 451 GTCAATGCCA TGGACACCAA TCCGCTNGCG GCAGACCCTG TGGTTGTGAT  
 15 501 CAAAGAAGCC GNCGANGATT TCAGACGANG TNGTCTGGTA TTGAGCCGTT  
 551 TGACCGAGCG TAAAAATCCAT GTGTGTAAGG CAGCTGGCGC AGACGTGCCG  
 601 TCTGAAATG CTGCCAACAT CGAAACACAT GAATTCGGCG GCCCGCATCC  
 651 GCGCGGTTTG AGTGGCACGC ACATTCATTT CATTGAGCCG GTCGGTGCAA  
 701 ACAAACCGT TTGGACCATC AATATCAAG ATGTAATTGC CATCGGACGT  
 20 751 TTGTTTGCAA CAGCCGCTCT GAACACCGAG CGCGTGATTG CTTTGGGTGG  
 801 TTCTCAAGTC AACAAACCAC GCCTCTTGCG TACCGTTTTG GGTGCGAAAG  
 851 TATCGCAAAT TACTGCGGGC GAATTGGTTG ACGCAGACAA CCGCGTGATT  
 901 TCCGGTTTCG TATTGAACGG CGCGATTACA CAAGGCGCGC ACGATTATTT  
 951 GGGACGCTAC CACAATCAGA TTTCCGTTAT CGAAGAAGGC CGCAGCAAAG  
 25 1001 AGCTGTTTCG CTGGGTTGCG CCGCAGCCGG ACAAATACTC CATCACGCGT  
 1051 ACGACCTCG GCCATTTCTT GAAAAACAAA CTCTTCAAGT TCACGACAGC  
 1101 CGTCAACGGT GCGGACCGCG CCATGGTGCC GATTGGTACT TACGAGCGCG  
 1151 TAATGCCGCT AGACATCCTG CCTACCCTGC TTTGCGCGA TTTAATCGTC  
 1201 GCGGATACCG ACAGCGCGCA AGCATTGGGT TGCTTGGAA TGGACGAAGA  
 30 1251 AGACCTCGCT TTGTGCAGCT TCGTCTGCCC GGGCAAATAC GAATANGGCC  
 1301 CGCTGTTGCG TAAGGTGCTG GAAACCNTTG AGAAGGAAGG CTGA

This encodes a protein having amino acid sequence <SEQ ID 128; ORF22a>:

35 1 MIKIKKGLNL PIAGRPEQVI YDGPVITEVA LLGEEYAGMR PXMVKKEGDA  
 51 VKKGQVLFED KXPGVVFTA PVSGKIAAIH RGEKRVLSV VIAVEGNDEI  
 101 EFERIYAPEAL ANLSGXEXXX NLIQSGLWTA LRXPFSKIP AVDAEPFAIF  
 151 VNAMDNTNPLA ADPVVVIKEA XXDFRXXLV LSRLTERKIH VCKAAGADVP  
 201 SENAANIETH EFGGPHFAGL SGTHIFIEP VGANKTVWTI NYQDVIAIGR  
 251 LFATGRLNTE RVIALGGSQV NKPRLLRVLV GAKVSQITAG ELVDADNRVI  
 301 SGSVLNGAIT QGAHDYLGRY HNQISVIEEG RSKELFGWVA POPDKYSITR  
 40 351 TTLGHFLKNK LFKETTAVNG GDRAMVPIGT YERVMPLDIL PTLLLRDLIV  
 401 GDTDSAQALG CLELDEEDLA LCSFVCPGKY EXGPLLRKVL ETXEKEG\*

The originally-identified partial strain B sequence (ORF22) shows 94.2% identity over a 158aa overlap with ORF22a:

45	orf22.pep	10	20	30	40	50	60
		MIKIKKGLNLPIAGRPEQAVYDGPVITEVALLGEEYAGMRPXMVKKEGDAVKKGQVLFED					
	orf22a	MIKIKKGLNLPIAGRPEQVIYDGPVITEVALLGEEYAGMRPXMVKKEGDAVKKGQVLFED					
		10	20	30	40	50	60
50	orf22.pep	70	80	90	100	110	120
		KKNPGVVFTAPASGKIAAIHRGEKRVLSVIAVEGNDEIEFERIYAPEALANLSGEEVRR					
	orf22a	KXPGVVFTAPVSGKIAAIHRGEKRVLSVIAVEGNDEIEFERIYAPEALANLSGXEXXX					
		70	80	90	100	110	120
55	orf22.pep	130	140	150			
		NLIQSGLWTALRTRPFSKIPAVDAEPFAIFVNAMDNTN					
	orf22a	NLIQSGLWTALRXPFSKIPAVDAEPFAIFVNAMDNTNPLAADPVVVIKEAXXDFRXXLV					
60		130	140	150	160	170	180

The complete strain B sequence (ORF22-1) and ORF22a show 94.9% identity in 447 aa overlap:

		10	20	30	40	50	60
	orf22a.pep	MIKIKKGLNLP	PIAGRPEQVI	YDGPVITEV	ALLGEEYAGMR	PXMKVKEGDA	VKKKGQVLFED
	orf22-1	MIKIKKGLNLP	PIAGRPEQAV	YDGPVITEV	ALLGEEYAGMR	PSMKVKEGDA	VKKKGQVLFED
5		10	20	30	40	50	60
	orf22a.pep	70	80	90	100	110	120
	orf22-1	KKXPGVVFTAP	VSGKIAAIHR	GEKRVLSV	VIAVEGNDEI	EFERYAPEAL	ANLSGXEXXX
10		70	80	90	100	110	120
	orf22a.pep	KKXPGVVFTAP	VSGKIAAIHR	GEKRVLSV	VIAVEGNDEI	EFERYAPEAL	ANLSGXEXXX
	orf22-1	KKXPGVVFTAP	VSGKIAAIHR	GEKRVLSV	VIAVEGNDEI	EFERYAPEAL	ANLSGXEXXX
15		130	140	150	160	170	180
	orf22a.pep	NLIQSGLWTA	LRTRPFSKIP	AVDAEPFAI	FVNAMDTNPL	AADPVVVIKE	AXDXFRXXLV
	orf22-1	NLIQSGLWTA	LRTRPFSKIP	AVDAEPFAI	FVNAMDTNPL	AADPTVVIKE	AAEDFKRGLLV
20		130	140	150	160	170	180
	orf22a.pep	190	200	210	220	230	240
	orf22-1	LSRLTERKIH	VCKAAGADV	PSENAANIET	HEFGGPHPAG	LSGTHIHFI	EPVVGANKTVWTI
25		190	200	210	220	230	240
	orf22a.pep	LSRLTERKIH	VCKAAGADV	PSENAANIET	HEFGGPHPAG	LSGTHIHFI	EPVVGANKTVWTI
	orf22-1	LSRLTERKIH	VCKAAGADV	PSENAANIET	HEFGGPHPAG	LSGTHIHFI	EPVVGANKTVWTI
30		250	260	270	280	290	300
	orf22a.pep	NYQDVIAIGR	LFGRLNTER	VIALGGSQV	NKPRLLRTVL	GAKVSQLTAG	ELVDADNRVI
	orf22-1	NYQDVITIGR	LFGRLNTER	VIALGGSQV	NKPRLLRTVL	GAKVSQLTAG	ELVDADNRVI
35		250	260	270	280	290	300
	orf22a.pep	310	320	330	340	350	360
	orf22-1	SGSVLNGAIT	QGAHDYLG	RYHNQISV	IEEGRSKEL	FGWVAPQPD	KYSITRRTLGHFLKNK
40		310	320	330	340	350	360
	orf22a.pep	SGSVLNGAIT	QGAHDYLG	RYHNQISV	IEEGRSKEL	FGWVAPQPD	KYSITRRTLGHFLKNK
	orf22-1	SGSVLNGAIT	QGAHDYLG	RYHNQISV	IEEGRSKEL	FGWVAPQPD	KYSITRRTLGHFLKNK
45		370	380	390	400	410	420
	orf22a.pep	LFKFTTAVNG	DRAMVPIGT	YERVMPLD	LPTLLRLD	LIVGDTDSA	QALGCLELDEEDLA
	orf22-1	LFKFTTAVNG	DRAMVPIGT	YERVMPLD	LPTLLRLD	LIVGDTDSA	QALGCLELDEEDLA
	orf22a.pep	370	380	390	400	410	420
	orf22-1	LFKFTTAVNG	DRAMVPIGT	YERVMPLD	LPTLLRLD	LIVGDTDSA	QALGCLELDEEDLA
	orf22a.pep	430	440				
	orf22-1	LCSFVCPGKY	EXGPLLRR	KVLETXE	KEGX		
	orf22a.pep	430	440				
	orf22-1	LCSFVCPGKY	EXGPLLRR	KVLETIE	KEGX		

Further work identified a partial gene sequence <SEQ ID 129> from *N.gonorrhoeae*, which encodes the following amino acid sequence <SEQ ID 130; ORF22ng>:

50	1	MIKIKKGLNL	PIAGRPEQVI	YDGPVITEV	ALLGEEYVGM	PSMKIKEGEA
	51	VKKKGQVLFED	KKXPGVVFTA	PASGKIAAIH	RGEKRVLSV	VIAVEGNDEI
	101	EFERYVPEAL	AKLSSEKVR	NLIQSGLWTA	LRTRPFSKIP	AVDAEPFAIF
	151	VNAMDTNPLA	ADPTVVIKEA	AEDFKRGLLV	LSRLTERKIH	VCKAAGADV
	201	SENAANIETH	EFGGPHPAGL	SGTHIHFI	EPVVGANKTV	WTI
55	251	LFVTGRLNTE	RVVALGGLQV	NKPRLLRTVL	GAKVSQLTAG	ELVDADNRVI
	301	SGSVLNGAIA	QGAHDYLG	RYHN*		

Further work identified complete gonococcal gene <SEQ ID 131>:

60	1	ATGATTAAAA	TCAAAAAAGG	TCTAAATCTG	CCCATCGCGG	GCAGACCGGA
	51	GCAAGTCATT	TATGACGGCC	CGGCCATTAC	CGAAGTCGCG	TTGCTGGCG
	101	AAGAATATGT	CGGCATGCGC	CCCTCGATGA	AAATCAAGGA	AGGTGAAGCC
	151	GTCAAAAAAG	GCCAAGTGCT	GTTGAAGAC	AAAAAGAATC	CGGCGTAGT
	201	ATTTACTGCG	CCGGCTTCAG	GCAAAATCGC	CGCTATTAC	CGTGGCGAAA
	251	AGCGCGTACT	TCAGTCAGTC	GTGATTGCGG	TTGAAGGCAA	CGACGAAATC
	301	GAGTTCGAAC	GCTACGTACC	TGAAGCGCTG	GCAAAATTGA	GCAGCGAAAA

5  
10  
15  
20

```

351 AGTGCGCCGC AACCTGATTC AATCAGGCTT ATGGACTGCG CTTGCGACCC
401 GTCCGTTTCAG CAAAATCCCT GCCGTAGATG CCGAGCCGTT CGCCATCTTC
451 GTCAATGCGA TGGACACCAA TCCGCTGGCT GCCGACCCTA CGGTCATCAT
501 CAAAGAAGCC GCCGAAGACT TCAAACGCGG CCTGTTGGTA TTGAGCCGCC
551 TGACCGAACG TAAAAATCCAT GTGTGTAAAG CAGCAGGCGC AGACGTGCCG
601 TCTGAAAATG CTGCCAATAT CGAAACACAT GAATTTGGCG GCCCGCATCC
651 TGCCGGCTTG AGTGGCACGC ACATTCATTT CATCGAGCCA GTCGGCGCGA
701 ATAAAACCGT GTGGACCATC AATTATCAAG ACGTGATTGC TATCGGACGT
751 TTGTTCGTAA CAGGCCGTCT GAATACCGAG CGCGTGGTTG CCTTGGGCGG
801 CCTGCAAGTC AACAAACGCG GCCTCTTGCG TACCGTTTGG GGTGCGAAGG
851 TGTCTCAACT TACCGCCGCG GAATTGGTTG ACGCGGACAA CCGCGTGATT
901 TCCGGTTCGG TATTGAACGG TGCGATTGCA CAAGCGCGCG ATGATTATTT
951 GGGACGCTAC CACAATCAGA TTTCCGTTAT CGAAGAAGGC CGCAGCAAAG
1001 AGCTGTTGCG CTGGGTTGCG CCGCAGCCGG ACAAATACTC CATCAGCGCG
1051 ACCACTCTCG GCCATTTCTT AAAAAACAAA CTCTTCAAGT TCACGACAGC
1101 CGTCAACGGC GGCGACCGCG CCATGTTACC GATCGGCACT TATGAGCGCG
1151 TAATGCCGTT GGACATCCTG CCTACCTTGC TTTTGCGCGA TTTAATCGTC
1201 GCGGATACCG ACAGCGCGCA GGCTTTGGGT TGCTTGAAT TGGACGAAGA
1251 AGACCTCGCT TTGTGCAGCT TCGTCTGCCC GGGCAAATAC GAATACGGCC
1301 CGCTGTTGCG CAAAGTGCTG GAAACCATTG AGAAGGAAGG CTGA

```

This encodes a protein having amino acid sequence <SEQ ID 132; ORF22ng-1>:

25  
30

```

1 MIKIKKGLNL PIAGRPEQVI YDGPATEVA LLGEEYVGM RPSMKIKEGEA
51 VKKGQVLFED KKNPGVVFTA PASGKIAAIH RGEKRVLSV VIAVEGNDEI
101 EFERYVPEAL AKLSSEKVR NLIQSLWTA LRTRPFSKIP AVDAEPFAIF
151 VNAMDNTPLA ADPTVIIKEA AEDFKRGLLV LSRLTERKIH VCKAAGADVP
201 SENAANIETH EFGGPHAGL SGTIHFIIEP VGANKTVWTI NYQDVIAIGR
251 LEVTGRLNTE RVVALGGLQV NKPRLLRTVL GAKVSQLTAG ELVDADNRVI
301 SGSVLNGAIA QGAHDYLG RYHNQISVIEE RSKELFGWVA PQPDKYSITR
351 TTLGHFLKNK LEKFTTAVNG GDRAMVPIGT YERVMLDIL PTLLLRDLIV
401 GDTDSAQALG CLELDEEDLA LCSFVCPGKY EYGPLLRKVL ETIEKEG*

```

The originally-identified partial strain B sequence (ORF22) shows 93.7% identity over a 158aa overlap with ORF22ng:

35  
40  
45

```

orf22.pep MIKIKKGLNLPIAGRPEQAVYDGPATEVALLGEEYAGMRPSMKVKEGDAVKKQVLFED 60
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf22ng MIKIKKGLNLPIAGRPEQVIYDGPATEVALLGEEYVGM RPSMKIKEGEAVKKQVLFED 60

orf22.pep KKNPGVVFTAPASGKIAAIHRGEKRVLSVVI AVEGNDEIEFERYAPEALANLSGEEVRR 120
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf22ng KKNPGVVFTAPASGKIAAIHRGEKRVLSVVI AVEGNDEIEFERYVPEALAKLSSEKVR 120

orf22.pep NLIQSLWTALRTRPFSKIPAVDAEPFAIFVNAMDNTN 158
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf22ng NLIQSLWTALRTRPFSKIPAVDAEPFAIFVNAMDNTNPLAADPTVIIKEAAEDFKRGLLV 180

```

The complete sequences from strain B (ORF22-1) and gonococcus (ORF22ng) show 96.2% identity in 447 aa overlap:

50  
55  
60

```

10 20 30 40 50 60
orf22-1.pep MIKIKKGLNLPIAGRPEQAVYDGPATEVALLGEEYAGMRPSMKVKEGDAVKKQVLFED
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf22ng-1 MIKIKKGLNLPIAGRPEQVIYDGPATEVALLGEEYVGM RPSMKIKEGEAVKKQVLFED
10 20 30 40 50 60

70 80 90 100 110 120
orf22-1.pep KKNPGVVFTAPASGKIAAIHRGEKRVLSVVI AVEGNDEIEFERYAPEALANLSGEEVRR
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf22ng-1 KKNPGVVFTAPASGKIAAIHRGEKRVLSVVI AVEGNDEIEFERYVPEALAKLSSEKVR
70 80 90 100 110 120

130 140 150 160 170 180
orf22-1.pep NLIQSLWTALRTRPFSKIPAVDAEPFAIFVNAMDNTNPLAADPTVIIKEAAEDFKRGLLV

```



5

10

15

20

25

30

35

ORF22 and this 48kDa protein show 72% aa identity in 158aa overlap:

40

45

orf22 121 NLIQSGLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNP 158  
NLI+SGLWTA RTRPFSK+PA+DA P +IFVNAMDTNP  
48kDa 121 NLIESGLWTAFRTRPFSKVPALDAIPSSIFVNAMDTNP 158

50

gi|1185395 (U24492) 48 kDa outer membrane protein [Actinobacillus pleuropneumoniae]  
Length = 449

Score = 530 bits (1351), Expect = e-150

Identities = 274/450 (60%), Positives = 323/450 (70%), Gaps = 4/450 (0%)

55

Query: 1 MIKIKKGLNLPIAGRPEQVIYDGPVITEVALLGEEYAGMRFXMKVKEGDAVKKGQVLFED 60  
MI IKKGL+LPIAG P QVI++G + EVA+LGEY GMRF MKV+EGD VKKGQVLFED  
Sbjct: 1 MITIKKGLDLPIAGTPAQVIHNGNTVNEVAMLGEEYVGMRPSMKVREGDVVKKGQVLFED 60

60

Query: 61 KKPVGVTAPVSGKIAAIHRGEKRVLSQSVVIAVEGNDIEFERYAPEALANLSGXEXX 120  
KK PGVVTAP SG + I+RGEKRVLSQSVI VEG+++I F RY LA+LS +  
Sbjct: 61 KKNPGVVTAPASGTVVTINRGEKRVLSQSVVIKVEGDEQITTRYEAAQLASLSAEQVKO 120

Query: 121 NLIQSGLWTALRXRPFSKIPAVDAEPFAIFVNAMDTNPLAADPVVVIKEAXXDFRRXXLV 180  
 NLI+SGLWTA R RPFSSK+PA+DA P +IFVNAMDTNPLAADP VV+KE DF+ V  
 Sbjct: 121 NLIESGLWTAFRTRPFSKVPALDAIPSSIFVNAMDTNPLAADPEVVLKEYETDFKDGLTV 180

5 Query: 181 LSRL--TERKIHVCKAAGADVP-SENAANIETHEFGGPHAGLSGTHIHFIIEPVGANKTV 237  
 L+RL ++ +++CK A +++P S I F G HPAGL GTHIHFI++PVGA K V  
 Sbjct: 181 LTRLFNGQKPVYLCKDADSNIPSPAIEGITIKSFSGVHPAGLVGTHIHFIHFDVPVGATKQV 240

10 Query: 238 WTINYQDVIAIGRLFATGRNLNTERVIALGGSQVKNPRLRLTVLGAKVSQITAGELVDADN 297  
 W +NYQDVIAIG+LF TG L T+R+I+L G QV PRL+RT LGA +SQ+TA EL +N  
 Sbjct: 241 WHLNYQDVIAIGKLFITTGELFTDRIISLAGPQVKNPRLVTRLGANLSQLTANELNAGEN 300

15 Query: 298 RVISGSLVNGAITQGAHDYLGRYHNQISVIEEGRSKELFGWVAPQDPKYSITRTTLGHFL 357  
 RVISGSLV+GA G DYLGRY Q+SV+ EGR KELFGW+ P DK+SITRT LGHF  
 Sbjct: 301 RVISGSLVSGATAAGPVDYLGRYALQVSVLAEGREKELFGWIMPGSDKFSITRTVLGHFG 360

20 Query: 358 KKNLKFKFTTAVNGGDRAMVPIGTYERVMXXXXXXXXXXXXXXXXXVGDTSAQXXXXXXXXXX 417  
 K KLF FTTAV+GG+RAMVPIG YERVM GDTDSAQ  
 Sbjct: 361 K-KLFNFTTAVHGGGERAMVPIGAYERVMPLDIPTLLLRDLAAGDTSAQNLGCLELDEE 419

Query: 418 XXXXXSFVCPGKYEXGPLLRLKVLLETXEKEG 447  
 ++VCPGK GP+LR LE EKEG

ORF22ng-1 also shows homology with the OMP from *A. pleuropneumoniae*:

gi|1185395 (U24492) 48 kDa outer membrane protein [Actinobacillus  
 pleuropneumoniae] Length = 449  
 Score = 555 bits (1414), Expect = e-157  
 Identities = 284/450 (63%), Positives = 337/450 (74%), Gaps = 4/450 (0%)

25 Query: 27 MIKIKKGLNLPPIAGRPEQVIYDGPATEVALLGEEYVGMPSMKIKEGEAVKKQGVLFED 86  
 MI IKKGL+LPIAG P QVI++G + EVA+LGEYVGMPSMK++EG+ VKKQGVLFED  
 Sbjct: 1 MITIKKGLDLPIAGTPAQVIHNGNTVNEVAMLGEEYVGMPSMKVREGDVVKKQGVLFED 60

30 Query: 87 KKNPGVVFTAPASGKIAAIHRGEKRVLSVVIIEGNDIEFEERYVPEALAKLSSEKVR 146  
 KKNPGVVFTAPASG + I+RGEKRVLSVVI VEG+++I F RY LA LS+E+V++  
 Sbjct: 61 KKNPGVVFTAPASGTVTINRGEKRVLSVVIKVEGDEQITFTRYEAAQLASLSAEQVKQ 120

35 Query: 147 NLIQSGLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNPLAADPTVIEKEAEDFKRGLLV 206  
 NLI+SGLWTA RTRPFSK+PA+DA P +IFVNAMDTNPLAADP V++KE DFK GL V  
 Sbjct: 121 NLIESGLWTAFRTRPFSKVPALDAIPSSIFVNAMDTNPLAADPEVVLKEYETDFKDGLTV 180

40 Query: 207 LSRL--TERKIHVCKAAGADVP-SENAANIETHEFGGPHAGLSGTHIHFIIEPVGANKTV 263  
 L+RL ++ +++CK A +++P S I F G HPAGL GTHIHFI++PVGA K V  
 Sbjct: 181 LTRLFNGQKPVYLCKDADSNIPSPAIEGITIKSFSGVHPAGLVGTHIHFIHFDVPVGATKQV 240

45 Query: 264 WTINYQDVIAIGRLFVTGRNLNTERVVALGGLQVKNPRLRLTVLGAKVSQITAGELVDADN 323  
 W +NYQDVIAIG+LF TG L T+R+++L G QV PRL+RT LGA +SQLTA EL +N  
 Sbjct: 241 WHLNYQDVIAIGKLFITTGELFTDRIISLAGPQVKNPRLVTRLGANLSQLTANELNAGEN 300

50 Query: 324 RVISGSLVNGAIAQGAHDYLGRYHNQISVIEEGRSKELFGWVAPQDPKYSITRTTLGHFL 383  
 RVISGSLV+GA A G DYLGRY Q+SV+ EGR KELFGW+ P DK+SITRT LGHF  
 Sbjct: 301 RVISGSLVSGATAAGPVDYLGRYALQVSVLAEGREKELFGWIMPGSDKFSITRTVLGHFG 360

55 Query: 384 KKNLKFKFTTAVNGGDRAMVPIGTYERVMXXXXXXXXXXXXXXXXXVGDTSAQXXXXXXXXXX 443  
 K KLF FTTAV+GG+RAMVPIG YERVM GDTDSAQ  
 Sbjct: 361 K-KLFNFTTAVHGGGERAMVPIGAYERVMPLDIPTLLLRDLAAGDTSAQNLGCLELDEE 419

60 Query: 444 XXXXXSFVCPGKYEYGPPLLRLKVLLETIEKEG 473  
 ++VCPGK YGP+LR LE IEKEG  
 Sbjct: 420 DLALCTYVCPGKNYGPMLRAALEKIEKEG 449

Based on this analysis, including the homology with the outer membrane protein of *Actinobacillus pleuropneumoniae*, it was predicted that these proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF22-1 (35.4kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 5A shows the results of affinity purification of the GST-fusion protein, and Figure 5B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result) and FACS analysis (Figure 5C). These experiments confirm that ORF22-1 is a surface-exposed protein, and that it is a useful immunogen.

### Example 16

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 133>:

```

10      1  ..GCGnCGnAAA TCATCCATCC CC..nACGTC GTAGGCCCTG AAGCCAACTG
      51  GTTTTTTATG GTAGCCAGTA CGTTTGTGAT TGCTTTGATT GGTTATTTTG
     101  TTACTGAAAA AATCGTCGAA CCGCAATTGG GCCCTTATCA ATCAGATTTG
     151  TCACAAGAAG AAAAAGACAT TCGGCATTCC AATGAAATCA CGCCTTTGGA
     201  ATATAAAGGA TTAATTTGGG CTGGCGTGGT GTTTGTGTC TTATCCGCCC
     251  TATTGGCTTG GAGCATCGTC CCTGCCGACG GTATTTGCG TCATCCTGAA
     301  ACAGGATTGG TTTCCGGTTC GCCGTTTTTA AAATCGATTG TTGTTTTTAT
     351  TTTCTTGTG TTTGCACTGC CGGGCATTGT TTATGGCCGG GTAACCCGAA
     401  GTTTGCGCGG CGAACAGGAA GTCGTTAATG CGmyGGCCGA ATCGATGAGT
     451  ACTCTGGsGC TTTmTTTGsw CAkcATCTTT TTTGCCGCAC AGTTTGTGCG
     501  ATTTTTTAAT TGGACGAATA TTGGGCAATA TATTGCCGTT AAAGGGCGCA
     551  CGTTCTTAAA AGAAGTCGGC TTGGGCGGCA GCGTGTTGTT TATCGGTTTT
     601  ATTTTAATTT GTGCTTTTAT CAATCTGATG ATAGGCTCCG CCTCCGCGCA
     651  ATGGGCGGTA ACTGCCCGCA TTTTCGTCCC TATGCTGATG TTGGCCGGCT
     701  ACGCGCCGA AGTCATCAA GCCGCTTACC GCATCGGTGA TTCCGTTACC
     751  AATATTATTA CGCCGATGAT GAGTTATTTT GGGCTGATTA TGGCGACGGT
     801  GrkCmmnTAC AAAAAAGATG CGGGCGTGGG TaCGcTGATT wCTATGATGT
     851  TGCCGTATTC CGCTTTCTTC TTGATTGCgT GGATTGCCTT ATTCTGCATT
     901  TGGGTATTTg TTTTGGGCCT GCCCGTCGGT CCCGGCGCGC CCACATCTA
     951  TCCCGCACCT TAA

```

This corresponds to the amino acid sequence <SEQ ID 134; ORF12>:

```

30      1  ..AXXIIHPXXV VGPEANWFFM VASTFVIALI GYFVTEKIVE PQLGPYQSDL
      51  SQEEKDIRHS NEITPLEYKG LIWAGVVFVA LSALLAWSIV PADGLLRHPE
     101  TGLVSGSPFL KSIVVFIFLL FALPGIVYGR VTRSLRGEQE VVNAXAESMS
     151  TLXLXLXXIF FAAQFVAFFN WTNIGQYIAV KGATFLKEVG LGGSVLFIFG
     201  ILICAFINLM IGSASAQWAV TAPIFVPLM LAGYAPEVIQ AAYRIGDSVT
     251  NIITPMMSYF GLIMATVXXY KKDAGVGTLI XMMLPYSAFF LIAWIALFCI
     301  WVFLGLPVG PGAPTFYPAP *

```

Further sequence analysis revealed the complete DNA sequence <SEQ ID 135> to be:

```

40      1  ATGAGTCAAA CCGATACGCA ACGGGACGGA CGATTTTAC GCACAGTCGA
      51  ATGGCTGGGC AATATGTTGC CGCATCCGGT TACGCTTTT ATTATTTTCA
     101  TTGTGTTATT GCTGATTGCC TCTGCCGTCG GTGCGTATTT CGGACTATCC
     151  GTCCCCGATC CGGCCCTGT TGGTGCGAAA GGACGTGCCG ATGACGGTTT
     201  GATTTACATT GTCAGCCTGC TCAATGCCGA CGGTTTTATC AAAATCCTGA
     251  CGCATACCGT TAAAAATTTC ACCGGTTTCG CGCGTTGGG AACGGTGTTG
     301  GTTTCTTTAT TGGGCGTGGG GATTGCGGAA AAATCGGGCT TGATTTCCGC
     351  ATTAATGCGC TTATTGCTCA CAAAATCGCC ACGCAAACCTC ACTACTTTTA
     401  TGGTTGTTTT TACAGGGATT TTATCTAATA CCGCTTCTGA ATTGGGCTAT
     451  GTCGTCTTAA TCCTTTTGTG CGCCATCATC TTTCATTCCC TCGGCCGCCA
     501  TCCGCTTGCC GGTCTGGCTG CGGCTTTCGC CGGCGTTTCG GCGGTTATT
     551  CGGCCAATCT GTTCTTAGGC ACAATCGATC CGCTCTTGGC AGGCATCACC
     601  CAACAGGCGG CGCAAATCAT CCATCCCGAC TACGTCGTAG GCCCTGAAGC
     651  CAACTGGTTT TTTATGGTAG CCAAGTACGT TGTGATTGCT TTGATTGGTT
     701  ATTTTGTAC TGAAAAATC GTCGAACCGC AATTGGGCC TATCAATCA
     751  GATTTGTCAC AAGAAGAAAA AGACATTCGG CATTCCAATG AAATCACGCC
     801  TTTGGAATAT AAAGGATTAA TTTGGGCTGG CGTGGTGTTT GTTGCCTTAT

```

5  
10  
15

```

851 CCGCCCTATT GGCTTGAGC ATCGTCCCTG CCGACGGTAT TTTGCGTCAT
901 CCTGAAACAG GATTGGTTTC CGGTTCCGCC TTTTAAAT CGATTGTTGT
951 TTTTATTTTC TTGTTGTTTG CACTGCCGGG CATTGTTTAT GGCCGGGTAA
1001 CCCGAAGTTT GCGCGGCGAA CAGGAAGTCG TTAATGCGAT GGCCGAATCG
1051 ATGAGTACTC TGGGGCTTTA TTTGGTCATC ATCTTTTGTG CCGCACAGTT
1101 TGTCGCATTT TTTAATTGGA CGAATATTGG GCAATATATT GCCGTTAAAG
1151 GGGCGACGTT CTTAAAAGAA GTCGGCTTGG GCGGCAGCGT GTGTTTATC
1201 GGTTTTATTT TAATTTGTGC TTTTATCAAT CTGATGATAG GCTCCGCCTC
1251 CCGCAATGG GCGGTAAGT GCGCGATTTT CGTCCCTATG CTGATGTTGG
1301 CCGGCTACGC GCGCAAGTC ATTCAAGCCG CTTACCGCAT CGGTGATTCC
1351 GTTACCAATA TTATTACGCC GATGATGAGT TATTTCCGGC TGATTATGGC
1401 GACGGTGATC AAATACAAA AAGATGCGGG CGTGGGTACG CTGATTCTA
1451 TGATGTTGCC GTATTCCGCT TTCTTCTTGA TTGCGTGGAT TGCCCTATTC
1501 TGCATTGGG TATTTGTTT GGGCCTGCCC GTCGGTCCCG GCGCGCCAC
1551 ATTCTATCCC GCACCTTAA

```

This corresponds to the amino acid sequence <SEQ ID 136; ORF12-1>:

20  
25

```

1 MSQTDTORRG RFLRTVEWLG NMLPHPVTLF IIFIVLLLIA SAVGAYFGLS
51 VPDPRPVGAK GRADDGLIYI VSLLNADGFI KILTHTVKNF TGFAPLGTVL
101 VSLLGVGIAE KSLGISALMR LLTKSPRKL TTFMVVFTGI LSNTASELGY
151 VVLIPLSAII FHSILGRHPLA GLAAAFAGVS GGYSANFLG TIDPLLAGIT
201 QQAQIIHPD YVVGPEANWF FMVASTFVIA LIGYFVTEKI VEPQLGPYQS
251 DLSQEEKDIR HSNEITPLEY KGLIWAGVVF VALSALLAWS IVPADGILRH
301 PETGLVSGSP FLKSIVVFIF LLFALPGIVY GRVTRSLRGE QEVVNAMAES
351 MSTLGLYLV IFFAAQFVAF FNWTNIGQYI AVKGATFLKE VGLGGSVLFI
401 GFILICAFIN LMIGSASAQW AVTAPIFVPM LMLAGYAPEV IQAAYRIGDS
451 VTNIITPMMS YFGLIMATVI KYKKGAGVGT LISMMLPYSA FFILIAWIALF
501 CIWVFEVLGLP VGPGAPTFYP AP*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

30 ORF12 shows 96.3% identity over a 320aa overlap with an ORF (ORF12a) from strain A of *N. meningitidis*:

35  
40  
45  
50  
55  
60

```

               10      20      30
orf12.pep      AXXIHPXXVVGPEANWFFMVASTFVIALI
               | | | | | | | | | | | | | | | | | | | | | | | | | | | |
orf12a      AAAFAGVSGGYSANFLGTIDPLLAGITQQAQIIHPDYVVGPEANWFFMVASTFVIALI
               180      190      200      210      220      230

               40      50      60      70      80      90
orf12.pep      GYFVTEKIVEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIV
               | | | | | | | | | | | | | | | | | | | | | | | | | | | |
orf12a      GYFVTEKIVEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIV
               240      250      260      270      280      290

               100      110      120      130      140      150
orf12.pep      PADGILRHPETGLVSGSPFLKSIVVFIFLLFALPGIVYGRVTRSLRGEQEVVNAXAESMS
               | | | | | | | | | | | | | | | | | | | | | | | | | | | |
orf12a      PADGILRHPETGLVSGSPFLKSIVVFIFLLFALPGIVYGRVTRSLRGEQEVVNAMAESMS
               300      310      320      330      340      350

               160      170      180      190      200      210
orf12.pep      TLXLXLXIIFFAAQFVAFFNWTNIGQYIAVKGATFLKEVGLGGSVLFIGFILICAFINLM
               | | | | | | | | | | | | | | | | | | | | | | | | | | | |
orf12a      TLGLYLVIIFFAAQFVAFFNWTNIGQYIAVKGATFLKEVGLGGSVLFIGFILICAFINLM
               360      370      380      390      400      410

               220      230      240      250      260      270
orf12.pep      IGSASAQWAVTAPIFVPMMLAGYAPEVIQAAYRIGDSVTNIITPMMSYFGLIMATVXXY
               | | | | | | | | | | | | | | | | | | | | | | | | | | | |
orf12a      IGSASAQWAVTAPIFVPMMLAGYAPEVIQAAYRIGDSVTNIITPMMSYFGLIMATVIKY
               420      430      440      450      460      470

               280      290      300      310      320

```

```
orf12.pep  KKDAGVGTLLXMMPLPYSAFFLIWIALFCIWVFLGLPVGPGAPTFFYPAPX
            |||
orf12a     KKDAGVGTLLSMMPLPYSAFFLIWIALFCIWVFLGLPVGPGAPTFFYPAPX
            480      490      500      510      520
```

5 The complete length ORF12a nucleotide sequence <SEQ ID 137> is:

```
1  ATGAGTCAAA CCGATACGCA ACGGGACGGA CGATTTTAC GCACAGTCGA
51  ATGGCTGGGC AATATGTTGC CGCACCCTGGT TACGCTTTT ATTATTTTCA
101 TTGTGTTATT GCTGATTGCC TCTGCCGCCG GTGCGTATT CGGACTATCC
151 GTCCCCGATC CGCGCCCTGT TGGTGCGAAA GGACGTGCCG ATGACGGTTT
10  201 GATTACAGTT GTCAGCGTGC TCGATGCTGA CGGTTTGATC AAAATCCTGA
251 CGCATACCGT TAAAAATTC ACCGGTTTCG CGCGTTGGG AACGGTGTG
301 GTTCTTTAT TGGGCGTGGG GATTGCGGAA AAATCGGGCT TGATTTCCGC
351 ATTAATGCGC TTATTGCTCA CAAAATCTCC ACGCAAATC ACTACTTTTA
401 TGGTTGTTT TACAGGGATT TTATCTAATA CCGCTTCTGA ATTGGGCTAT
15  451 GTCGTCCTAA TCCCTTTGTC CGCCATCATC TTTCATTCCC TCGCCGCCA
501 TCCGCTTGCC GGTCTGGCTG CGGCTTTCGC CGGCGTTTCG GCGGTTATT
551 CGGCCAATCT GTCTTAGGC ACAATCGATC CGCTCTTGGC AGGCATCACC
601 CAACAGGCGG CGCAAATCAT CCATCCCGAC TACGTCGTAG GCCCTGAAGC
651 CAACTGGTTC TTATGGTAG CAGTACGTT TGTGATTGCT TTGATTGGTT
20  701 ATTTTGTAC TGAAGAAATC GTCGAACCGC AATTGGGCCC TTATCAATCA
751 GATTGTGAC AAGAAGAAAA AGACATTCCA CATTCCAATG AAATCAGGCC
801 TTTGGAATAT AAAGGATTAA TTTGGGCTGG CGTGGTGTTC GTTGCCTTAT
851 CGGCCCTATT GGCTTGGAGC ATCGTCCCTG CCGACGGTAT TTTGCGTCAT
901 CCTGAAACAG GATTGGTTTC CGGTCGCCG TTTTAAAT CAATGTGTTG
25  951 TTTATTTC TTGTGTTTG CACTGCCGGG CATTGTTTAT GGC CGGGTAA
1001 CCGAAGTTT GCGCGCGCAA CAGGAAGTCG TTAATGCGAT GCGCGAATCG
1051 ATGAGTACTC TGGGGCTTTA TTTGGTCATC ATCTTTTTTG CCGCACAGTT
1101 TGTGCGATTT TTTAATTGGA CGAATATTGG GCAATATATT GCGGTTAAG
1151 GGGCGACGTT CTAAAAGAA GTCGGCTTGG GCGGCAGCGT GTTGTTCATC
30  1201 GGTTCATTT TAATTTGTGC TTTTATCAAT CTGATGATAG GCTCCGCCTC
1251 CGCGCAATGG GCGGTAATG CGCCGATTTT CGTCCCTATG CTGATGTTGG
1301 CCGGCTACGC GCCCGAAGTC ATTCAAGCCG CTTACCGCAT CGGTGATTCC
1351 GTTACCAATA TTATTACGCC GATGATGAGT TATTTCGGGC TGATTATGGC
1401 GACGGTGATC AAATACAAAA AAGATGCGGG CGTGGGTACG CTGATTTCTA
35  1451 TGATGTTGCC GTATTCGGCT TTCTTCTGA TTGCGTGGAT TGCCTTATTC
1501 TGCATTTGGG TATTTGTTTT GGGCTGCCC GTCGGTCCCG GCGCGCCAC
1551 ATTCTATCCC GCACCTTAA
```

This encodes a protein having amino acid sequence <SEQ ID 138>:

```
1  MSQTDQTDG RFLRTVEWLG NMLPHVTLF IIFIVLLLIA SAAGAYFGLS
40  51  VPDPRPVGAK GRADDGLIHV VSLLDADGLI KILTHTVKNF TGFAPLGTVL
101  VSLGVGIAE KSLISALMR LLLTKSPRKL TTFMVVFTGI LSNASELGY
151  VVLIPLSAII FHSLSRHLPLA GLAAAFAGVS GGYSANFLG TIDPLLAGIT
201  QQAAQIIHPD YVVGPEANWF FMVASTFVIA LIGYFVTEKI VEPQLGPYQS
45  251  DLSQEEKDIR HSNEITPLEY KGLIWAGVVF VALSALLAWS IVPADGILRH
301  PETGLVSGSP FLKSIVVFIF LLFALPGIVY GRVTRSLRGE QEVVNAMAES
351  MSTLGLYLV IFFAAQFVAF FNWTNIGQYI AVKGATFLKE VGLGGSVLFI
401  GFILICAFIN LMIGSASAQW AVTAPIFVPM LMLAGYAPEV IQAAYRIGDS
451  VTNIITPMMS YFGLIMATVI KYKKDAGVGT LISMMPLPYSA FFLIAWIALF
50  501  CIWVFLGLP VGPGAPTFFP AP*
```

ORF12a and ORF12-1 show 99.0% identity in 522 aa overlap:

```
10      20      30      40      50      60
orf12a.pep  MSQTDQTDGRFLRTVEWLG NMLPHVTLFI IIFIVLLLIASAAGAYFGLSVDPDRPVGAK
55  orf12-1   MSQTDQTDGRFLRTVEWLG NMLPHVTLFI IIFIVLLLIASAVGAYFGLSVDPDRPVGAK
10      20      30      40      50      60

70      80      90      100     110     120
orf12a.pep  GRADDGLIHVVSLLDADGLIKILTHTVKNFTGFAPLGTVLVSLGVGIAEKSLISALMR
60  orf12-1   GRADDGLIYVSLLDADGLIKILTHTVKNFTGFAPLGTVLVSLGVGIAEKSLISALMR
70      80      90      100     110     120
```

		130	140	150	160	170	180
	orf12a.pep	LLLT	KSPR	KLTT	FMV	VFTG	ILSNT
	orf12-1	LLLT	KSPR	KLTT	FMV	VFTG	ILSNT
5		130	140	150	160	170	180
	orf12a.pep	GGYS	ANLFL	GTID	PPLA	GITQ	QAAQ
10	orf12-1	GGYS	ANLFL	GTID	PPLA	GITQ	QAAQ
		190	200	210	220	230	240
	orf12a.pep	GGYS	ANLFL	GTID	PPLA	GITQ	QAAQ
15	orf12-1	GGYS	ANLFL	GTID	PPLA	GITQ	QAAQ
		250	260	270	280	290	300
	orf12a.pep	VEPQ	LGPY	QSDLS	QEEK	DIRH	SNEIT
20	orf12-1	VEPQ	LGPY	QSDLS	QEEK	DIRH	SNEIT
		310	320	330	340	350	360
	orf12a.pep	PETG	LVSG	SPFL	KSIV	VFI	FLF
25	orf12-1	PETG	LVSG	SPFL	KSIV	VFI	FLF
		370	380	390	400	410	420
	orf12a.pep	IFFA	AQFV	AFNWT	NIGQ	YIAV	KGAT
30	orf12-1	IFFA	AQFV	AFNWT	NIGQ	YIAV	KGAT
		430	440	450	460	470	480
	orf12a.pep	AVTA	PIFV	PMLM	LAGYA	PEVI	QAAY
35	orf12-1	AVTA	PIFV	PMLM	LAGYA	PEVI	QAAY
		490	500	510	520		
	orf12a.pep	LISM	MLPY	SAFF	LIAW	IALFC	IWV
40	orf12-1	LISM	MLPY	SAFF	LIAW	IALFC	IWV

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF12 shows 92.5% identity over a 320aa overlap with a predicted ORF (ORF12.ng) from *N.*

### 45 *gonorrhoeae*:

	orf12.pep	AXXI	IHPX	XVVG	PEAN	WFFM	VAST	FVIALI	30
	orf12ng	AAAF	AGVSG	GYSAN	LFLGT	IDP	LLAG	ITQQAQ	232
50	orf12.pep	GYFV	TEKIVE	PQLGP	YQSD	LSQEEK	DIRHS	NEITPLEY	90
	orf12ng	GYFV	TEKIVE	PQLGP	YQSD	LSQEEK	DIRHS	NEITPLEY	292
55	orf12.pep	PADG	ILRHP	PETGL	VSGSP	FLKSIV	VFI	FLF	150
	orf12ng	PADG	ILRHP	PETGL	VAGSP	FLKSIV	VFI	FLF	352
60	orf12.pep	TLXL	LXXI	FFAA	QFVA	FFNWT	NIGQ	YIAV	210
	orf12ng	TLGL	YLVII	FFAA	QFVA	FFNWT	NIGQ	YIAV	412
	orf12.pep	IGSAS	AQWAV	TAPI	FVPM	LMLAG	YAPE	VIQAAY	270
65	orf12ng	IGSAS	AQWAV	TAPI	FVPM	LMLAG	NAPQ	VIQAAY	472

```

orf12.pep      KKDAGVGTLIXMMLPYSAFFLIAWIALFCIWVFLGLPVGPGAPTFFYPAP  320
               |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf12ng        KKDAGVGTLISMMLPYSAFFLIAWIALFCIWVFLGLPVGPGTPTFFYPVP  522

```

The complete length ORF12ng nucleotide sequence <SEQ ID 139> is:

```

5      1  ATGAGTCAAA CCGACGCGCG TCGTAGCGGA CGATTTTAC GCACAGTCGA
      51  ATGGCTGGGC AATATGTTGC CGCACCCGGT TACGCTTTT ATTATTTTCA
     101  TTGTGTTATT GCTGATTGcc tctgCCGTCG GTGCGTATT CGGACTATCC
     151  GTCCCCGATC CGCGTCCTGT TGGGGCGAAA GGACGTGCCG ATGACGGTTT
     201  GATTACAGTT GTCAGCTGTC TCGATGCCGA CGGTTTGATC AAAATCCTGA
     251  CGCATACCGT TAAAAATTTC ACCGGTTTCG CGCGTTGGG AACGGTGTG
     301  GTTCTTTTAT TGGGCGTGGG GATTGCGGAA AAATCGGGCT TGATTTCCGC
     351  ATTAATGCGC TTATTGCTCA CAAAATCCCC ACGCAAACCTC ACTACTTTTA
     401  TGGTTGTTT TACAGGATT TTATCCAATA CGGCTTCTGA ATTGGGCTAT
     451  GTCGTCCTAA TCCCTTTGTC CGCCGTCATC TTTCATTGCG TCGGCGGCCA
     501  TCCGCTTGCC GGTGTTGGTG CGGCTTTTCG CGGCGTTTCG GCGGTTATT
     551  CGGCCAATCT GTTCTTAGGC ACAATCGATC CGCTCTTGGC AGGCATCACC
     601  CAACAGCGCG CGCAAATCAT CCATCCCGAC TACGTCGTAG GCCCTGAAGC
     651  CAACTGGTTT TTTATGGCAG CCAGTACGTT TGTGATTGCT TTGATTGGTT
     701  ATTTTGTAC TGAATAATC GTCGAACC GC AATGGGCC TTATCAATCA
     751  GATTGTGAC AAGAAGAAA AGACATTCG CATCCAATG AAATCAGCCG
     801  TTTGGAATAT AAAGGATTAA TTTGGGAGG CGTGGTGTG GTTGCTTAT
     851  CCGCCCTATT GGCTTGGAGC ATCGTCCCTG CCGACGGTAT TTTGCGTCAT
     901  CCTGAAACAG GATTGGTTGC CGGTTCCCGG TTTTAAAAT CGATTGTTGT
     951  TTTTATTTT TGTGTTTTCG CGTGGCGGG CATGTTTAT GGCCGGATAA
    1001  CCCGAAGTTT GCGCGGCGAA CGGGAAGTCG TTAATGCGAT GGCCGAATCG
    1051  ATGAGTACTT TGGGACTTTA TTTGGTCATC ATCTTTTTT CCGCACAGTT
    1101  TGTCGCATTT TTTAATTGGA CGAATATTGG GCAATATATT GCCGTTAAAG
    1151  GGGCGGTGTT CTTAAAGAA GTCGGCTTGG GCGGCAGTGT GTTGTATTATC
    1201  GGTTTTATTT TAATTTGTGC TTTTATCAAT CTGATGATAG GCTCCGCCTC
    1251  CGCGCAATGG GCGGTAATG CGCCGATTTT CGTCCCTATG CTGATGTTGG
    1301  CCGGCTACGC CCGGGAAGTC ATTCAAGCCG CTTACCGCAT CGGTGATTCC
    1351  GTTACCAATA TTATTACGCC GATGATGAGT TATTTCGGGC TGATTATGGC
    1401  GACGGTAATC AAATACAAA AAGATGCGGG CGTAGGCACG CTGATTTCTA
    1451  TGATGTTGCC GTATTCCGCT TTCTTCTTAA TTGCATGGAT CGCCTTATTC
    1501  TGCATTTGGG TATTGTTT GGTCTGCCC GTCGGTCCCG GCACACCCAC
    1551  ATTCTATCCG GTGCCTTAA

```

This encodes a protein having amino acid sequence <SEQ ID 140>:

```

      1  MSQTDARRSG RFLRTVEWLG NMLPHPVTLF IIFIVLLLIA SAVGAYFGLS
     51  VPDPRPVGAK GRADDGLIHV VSLLDADGLI KILTHTVKNF TGFAPLGTVL
    101  VSLLGVGIAE KSLGISALMR LLTKSPRKL TTFMVVFTGI LSNTASELG
    151  VVLIPLSAVI FHSLGRHPLA GLAAAFAGVS GGYSANLFLG TIDPLLAGIT
    201  QAAQIIHPD YVVGPEANWF FMAASTFVIA LIGYFVTEKI VEPQLGPYQS
    251  DLSQEEKDIR HSNEITPLEY KGLIWAGVVF VALSALLAWS IVPADGILRH
    301  PETGLVAGSP FLKSIVVFIF LLFALPGIVY GRITRSLRGE REVVNAMAES
    351  MSTLGLYLV IFFAAQFVAF FNNWNIGQYI AVKGAVFLKK FRLGGSVLFI
    401  GFILICAFIN LMIGSASAQW AVTAPIFVPM LMLAGNAPOV IQAAYRIGDS
    451  VTNIITPMS YFGLIMATVI KYKKGAVGT LISMMLPYSA FFLIAWIALF
    501  CIWVFLGLP VPGTPTFFYP VP*

```

ORF12ng shows 97.1% identity in 522 aa overlap with ORF12-1:

```

50      10      20      30      40      50      60
orf12-1.pep  MSQTDQTDGRFLRTVEWLG NMLPHPVTLF IIFIVLLLIASAVGAYFGLSVDPDRPVGAK
               |||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf12ng      MSQTDARRSGRFLRTVEWLG NMLPHPVTLF IIFIVLLLIASAVGAYFGLSVDPDRPVGAK
               10      20      30      40      50      60

55      70      80      90      100     110     120
orf12-1.pep  GRADDGLIYIVSLLNADGFIKILTHTVKNFTGFAPLGTVLVSLLGVGIAEKSLGISALMR
               |||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf12ng      GRADDGLIHVVSLLDADGLIKILTHTVKNFTGFAPLGTVLVSLLGVGIAEKSLGISALMR
               70      80      90      100     110     120

60      130     140     150     160     170     180
orf12-1.pep  LLTKSPRKLTTFMVVFTG LLSNTASELG YVVLIPLSAIIFHSLGRHPLAGLAAAFAGVS

```

	orf12ng	 LLLTSPKRLTTFMVVFTGILSNTASELGYVVLIPLSAVIFHSLGRHPLAGLAAAFAGVS 130 140 150 160 170 180
5	orf12-1.pep	190 200 210 220 230 240 GGYSANLFLGTIDPELLAGITQQAQIIHPDYVVGPEANWFFMVASTFVIALIGYFVTEKI 
10	orf12ng	190 200 210 220 230 240 GGYSANLFLGTIDPELLAGITQQAQIIHPDYVVGPEANWFFMAASTFVIALIGYFVTEKI 
	orf12-1.pep	250 260 270 280 290 300 VEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVVVALSALLAWSIVPADGILRH 
15	orf12ng	250 260 270 280 290 300 VEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVVVALSALLAWSIVPADGILRH 
	orf12-1.pep	310 320 330 340 350 360 PETGLVSGSPFLKSIVVFIFLLFALPGIVYGRVTRSLRGEQEVVNMAESMSTLGLYLVI 
20	orf12ng	310 320 330 340 350 360 PETGLVAGSPFLKSIVVFIFLLFALPGIVYGRITRSLRGEREVVNMAESMSTLGLYLVI 
	orf12-1.pep	370 380 390 400 410 420 IFFAAQFVAFNWTNIGQYIAVKGATFLKEVGLGGSVLFIFIGFILICAFINLMIGSASAQW 
25	orf12ng	370 380 390 400 410 420 IFFAAQFVAFNWTNIGQYIAVKGAVFLKEVGLGGSVLFIFIGFILICAFINLMIGSASAQW 
	orf12-1.pep	430 440 450 460 470 480 AVTAPIFVPMMLLAGYAPEVIAAYRIGDSVTNIITPMMSYFGLIMATVIKYKKDAGVGT 
30	orf12ng	430 440 450 460 470 480 AVTAPIFVPMMLLAGYAPEVIAAYRIGDSVTNIITPMMSYFGLIMATVIKYKKDAGVGT 
	orf12-1.pep	490 500 510 520 LISMMLPYSAFFLIAWIALFCIWFVFLGLPVGPGAPTFFYPAPX 
35	orf12ng	490 500 510 520 LISMMLPYSAFFLIAWIALFCIWFVFLGLPVGPGTPTFFYPVPX 

40 In addition, ORF12ng shows significant homology with a hypotehtical protein from *E.coli*:

sp|P46133|YDAH\_ECOLI HYPOTHETICAL 55.1 KD PROTEIN IN OGT-DBPA INTERGENIC REGION  
>gi|1787597 (AE000231) hypothetical protein in ogt 5' region [Escherichia coli]  
Length = 510  
Score = 329 bits (835), Expect = 2e-89  
Identities = 178/507 (35%), Positives = 281/507 (55%), Gaps = 15/507 (2%)

45 Query: 8 RSGRFLRTVEWLGNNMLPHPVTVXXXXXXXXXXASAVGAYFGLSVDPDRPVGAKGRADDGL 67  
+SG+ VE +GN +PHP +A+ + FG+S +P D  
Sbjct: 13 QSGKLYGWVERIGNKVPHFLLFIYLIIVLMVTTAILSAFGVSAKNP-----TDGTP 64

50 Query: 68 IHVVSLLDADGLIKILHTVKNFTGFAPXXXXXXXXXXIAEKSGLISALMRLLLTKSP 127  
+ V +LL +GL L + +KNF+GFAP +AE+ GL+ ALM + +  
Sbjct: 65 VVKNLLSV EGLHWFLPNVIKNFSGFAPLGAILALVLGAGLAERVGLLPALMVKMASHVN 124

55 Query: 128 RKLTTFMVVFTGILSNTASELGYVVLIPLSAVIFHSLGRHPLAGLAAAFAGVSGGYSANL 187  
+ ++MV+F S+ +S+ V++ P+ A+IF ++GRHP+AGL AA AGV G++ANL  
Sbjct: 125 ARYASYMVLFIAFFSHISSDAALVIMPPMGALIFLAVGRHPVAGLLAAIAGVCGFTANL 184

60 Query: 188 FLGTIDPELLAGITQQAQIIHPDYVVGPEANWFFMAASTFVIALIGYFVTEKIVEPQLGP 247  
+ T D LL+GI+ +AA +P V NW+FMA+S V+ ++G +T+KI+EP+LG  
Sbjct: 185 LIVTTDVLLSGISTEAAAAFNQMHVSVIDNWFMASSVVVLTIVGGLITDKIIEPRLGQ 244

65 Query: 248 YQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVVVALSALLAWSIVPADGILRHPETGLVA 307  
+Q + ++ + S GL AGVV + A +A ++P +GILR P V  
Sbjct: 245 WQGNSEKQLQTLTESQRF-----GLRIAGVVSLLFIAAIALMVIPQNGILRDPINHTVM 298

Query: 308 GSPFLKSIVVFIFLLFALPGIVYGRITRSLRGEREVVNMAESMSTLGLYLXXXXXXXXX 367  
SPF+K IV I L F + + YG TR++R + ++ + M E M + ++



Sbjct: 299 PSPFIKGIVPLIILFFVVSLEYGIATRTIRQADLPHLMIEPMKEMAGFIVMVFLAQF 358  
 Query: 368 XXXXNWTNIGQYIAVKAVFLKEVGLGGSVLFIFIGILICAFINLMIGSASAQWAVTAPIF 427  
 NW+N+G++IAV L+ GL G F+G L+ +F+ + I S SA W++ APIF  
 5 Sbjct: 359 VAMFNWSNMGKFIAVGLTDILESSGLSGIPAFVGLALLSSFLCMFIASGSAIWSILAPIF 418  
 Query: 428 VPMLMLAGYAPEVIQAAAYRIGDSVTNIITPMSYFGLIMATVIKYKKDAGVGTLSMMLP 487  
 VPM ML G+ P Q +RI DS + P+ + L + + +YK DA +GT S++LP  
 10 Sbjct: 419 VPMFMLLGHFPAFAQILFRIADSSVLPLAPVSPFVPLFLGLQRYKPKDAKLGTYYSVLVP 478  
 Query: 488 YSAFFLIAWIALFCIWWFVLGLPVGPG 514  
 Y FL+ W+ + W +++GLP+GPG  
 Sbjct: 479 YPLIFLVVWLLMLLAW-YLVGLPIGPG 504

- 15 Based on this analysis, including the presence of several putative transmembrane domains and the predicted actinin-type actin-binding domain signature (shown in bold) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 17

- 20 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 141>:

1 ..ACAGCCGGCG CAGCAGGTTn CnCGGTCTTC GTTTTCGTAA CGGACAGTCA  
 51 GGTGGAGGTG TTCGGGAACA TCCAGACCGC AGTGGAACA GGTTTTTTTC  
 101 ATGGCATTTC GGTTTCGTCT GTGTTTGGTG CGGCGGCACA AGACTCGGCA  
 151 ATgGCTTCGC GCAGTGCCTC TATACCGGTA TTTTCAGCAA CGGAAATGCG  
 25 201 GACGGcGgCA ATTTTCCCG CAGCGTCGCG CCATATGCCG GTGTTTTgTT  
 251 CTTCAGACGG CAGCAGGTCG GTTTTGTGT ACACCTTgAT GCACGGAaTA  
 301 TCGCCGGCAT GGATTCTTG CAGTACGTTT TCCACGTCTT CAATCTGCTG  
 351 TCCGCTGTTC GGAGCGGCGG CATCGACGAC GTGCAGCAGC ACATCgGcTT  
 401 gCGCGGTTTC TTCCAGCGTG GCgGAAAAGG CGGAAATCAG TTTgTGCGGC  
 30 451 agATyGCTnA CGAATCCGAC GGTATCGGTC AGGATAATGC TGCATTCCGG  
 501 ACT..

This corresponds to the amino acid sequence <SEQ ID 142; ORF14>:

1 ..TAGAAGXXVF VEVTDQSVEV FGNIQTAVET GFFHGISVSS VFGAAQDSA  
 51 MASRSASIPV FSATEMRTAA IFPAASRHMP VFCSSDGSRS VLLYTLMHGI  
 35 101 SPAWISCSTF STSSICCP LF GAAASTCSS TSACAVSSSV AEKAEISLCG  
 151 RXLTNPTVSV RIMLHSG..

Computer analysis of this amino acid sequence gave the following results:

### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF14 shows 94.0% identity over a 167aa overlap with an ORF (ORF14a) from strain A of *N.meningitidis*:

40 orf14.pep TAGAAGXXVFVFVTDQSVEVFGNIQTAVET  
 10 20 30  
 orf14a GRQLGFLRVGGALFVITAQARVNNALCDCLTTGAAGFAVFVTDGQMVFVGNVQPAVET  
 45 150 160 170 180 190 200  
 orf14.pep GFFHGISVSSVFGAAQDSAMASRSASIPVFSATEMRTAAIFPAASRHMPVFCSSDGSRS  
 50 orf14a GFFHGISVSSVFGAAQYSAMASRSASIPVFSATEMRTAAIFPAASRHMPVFCSSDGSRS  
 210 220 230 240 250 260

		100	110	120	130	140	150
orf14.pep	VLLYTLMHGISPAWISCSTFSTSSICPLFGAAASTTCSSTSACAVSSSSVAEKAEISLCG						
orf14a	VLLYTLMHGISPAWISCSTFSTSSICPLFGAAASTTCSSTSACAVSSSSVAEKAEISLCG						
5		270	280	290	300	310	320
			160				
orf14.pep	RSLTNPTVSVRIMLHSG						
10	orf14a	RSLTNPTVSVRIMLHSGLMYSRRRAVSSVAKSWSFAYMPDLVSRNLRLDPLTVX					
		330	340	350	360	370	380

The complete length ORF14a nucleotide sequence <SEQ ID 143> is:

	1	ATGGAGGATT	TGCAGGAAAT	CGGGTTCGAT	GTCGCCGCCG	TAAAGGTAGG
	51	TCGGCAGCGC	GAACATCATC	GTCTGCATCA	TCCCCAGCCC	GGCAACGGCG
15	101	AGGCGGACGA	TGTATTGTTT	GCGTTCCTTT	TGGTTGGCGG	CTTCGATTTT
	151	TTGCGCGTCA	TAGGGTGCGG	CGGTGTAGCC	TATCTGCCTG	ATTTTCAACA
	201	GAATGTCGGA	AAGGCGGATT	TTGCCGTCGT	CCCAGACGAC	GCGGCAGCGG
	251	TGCGTGCTGT	AATTGAGGTC	GATGCGGACG	ATGCCGTCGT	TACGCAAAAG
	301	CTGCTGTTCG	ATCAGCCAGA	CGCAGGCGGC	GCAGGTGATG	CCGCCGAGCA
20	351	TTAAAACCGC	CTCGCGCGTG	CCGCCGTGGG	TTTCCACAAA	GTCGGACTGG
	401	ACTTCGGGCA	GGTCGTACAG	GCGGATTTGG	TCGAGGATTT	CTTGGGGCGG
	451	CAGCTCGGTT	TTTTGCGCGT	CGCGCGTGCG	TTGTTTGTA	TAATGCCCCA
	501	AGCCCGCGTC	AATAATGCTT	TGTGCGACTG	CCTGACAACC	GGCGCAGCAG
	551	GTTTCGCGGT	CTTCGTTTTT	GTAACGGACG	GTCAGATGCA	GGTTTTCGGG
25	601	AACGTCCAGC	CCGCGTGGGA	AACAGGTTTT	TTTCATGGCA	TTTCGGTTTT
	651	GTCTGTGTTT	GGTGCGGCGG	CACAATACTC	GGCAATGGCT	TCGCGCAGTG
	701	CGTCTATACC	GGTATTTTCA	GCAACGGAAA	TGCGGACGGC	GGCAATTTTT
	751	CCCGCAGCGT	CGCGCCATAT	GCCCGTGTTC	TGTTCTTCAG	ACGGCAGCAG
	801	GTCCGTTTTG	TTGTACACCT	TGATGCACGG	AATATCGCCG	GCATGGATTT
30	851	CTTGCACTAC	GTTTTCCACG	TCTTCAATCT	GCTGTCCGCT	GTTTCGGAGCG
	901	GCGGCATCGA	CGAGTGCAG	CAGCACATCG	GCTTGCGCGG	TTTCTTCCAG
	951	CGTGCGGAAA	AAGGCGGAAA	TCAGTTTGTG	CGGCAGATCG	CTGACGAATC
	1001	CGACGGTATC	GGTCAGGATA	ATGCTGCATT	CGGGACTGAT	GTACAGCCGC
	1051	CGCGCCGTCG	TGTCGAGTGT	GGCGAAAAGC	TGGTCTTTCG	CATATATGCC
35	1101	CGACTTGGTC	AGCCGGTTGA	ACAGACTGGA	TTTGCCGACA	TTGGTATAG

This encodes a protein having amino acid sequence <SEQ ID 144>:

	1	MEDLQEI GFD	VAAVKVGRQR	EHHRLHHPQ	GNGEADDVLF	AFFLVGGFDF
	51	LRVIGCGGVA	YLPDFQQNVG	KADFAVVPDD	AAAVRAVIEV	DADDAVCTQK
	101	LLFDQPDAGG	AGDAAEH*NR	LARAAGVGHK	VGLDFGQVVQ	ADLVEDFLGR
40	151	QLGFLRVGGA	LFVITAQARV	NNALCDCLTT	GAAGFAVFVF	VTGQMVFVFG
	201	NVQPAVETGF	FHGISVSSVF	GAAQYSAMA	SRSASIPVFS	ATEMRTAAIF
	251	PAASRHPVVF	CSSDGSRSVL	LYTLMHGISP	AWISCSTFST	SSICPLFGA
	301	AASTTCSSTS	ACAVSSSSVAE	KAEISLCGRS	LTNPTVSVRI	MLHSGLMYSR
	351	RAVVSSVAKS	WSFAYMPDLV	SRLNRLDLPT	LV*	

45 It should be noted that this sequence includes a stop codon at position 118.

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF14 shows 89.8% identity over a 167aa overlap with a predicted ORF (ORF14.ng) from *N.gonorrhoeae*:

	orf14.pep	TAGAAGXXVFVFTDSQVEVFGNIQTAVET	30
50	orf14ng	GRQFGFFRVGGASFVITAQAGIDDALCDCLTADAAGFAVFAFVADGQMVFVGNVQPAVET	208
	orf14.pep	GFFHGISVSSVFGAAAQDSAMASRSASIPVFSATEMRTAAIFPAASRHPVFCSSDGSRS	90
55	orf14ng	GFFHGISVSSVFGAAAQYSAMASRSASIPVFSATEMRTAAIFPAASRHPVFCSSDGSRS	268
	orf14.pep	VLLYTLMHGISPAWISCSTFSTSSICPLFGAAASTTCSSTSACAVSSSSVAEKAEISLCG	150
60	orf14ng	VLLYTLMHGISWAWISCSTFSTSSICPLFRAAASTTCSSTSACTVSSKVAEKAEISLCG	328

orf14.pep RXLTNPTVSVRIMLHSG 167  
 | | | | | | | | | | | | | | | | |  
 orf14ng RSLTNPTVSVRIMLHAGLMYSRRRAVVSRAKSWSFAYMPDLVSRLNRLDLPTLV 382

The complete length ORF14ng nucleotide sequence <SEQ ID 145> is predicted to encode a protein

5 having amino acid sequence <SEQ ID 146>:

1 MEDLQEIGFD VAAVKVGRQR EHHRLHHTQS GNGKADDVLF AFFLVGGFDF  
 51 LRVIGCGGVA CLPDFQQNVG EADFAVPDD AAARAVIEV DADDAVCAQK  
 101 LLFDQPDAGG AGNAAEHQHC FVRAIMGFHK VGLDFGQVVQ ADLVEDFLGR  
 151 QFGFFRVGGA SFVITAQAGI DDALCDCLTA DAAGFAVFAF VADGQMQVFG  
 10 NVQPAVETGF FHGISVSSVF GAAQYSAMA SRSASIPVFS ATEMRTAAIF  
 201 PAASRHPVF CSSDGSRSVL LYTLMHGISW AWISCSTFST SSICCPFLRA  
 251 AASTTCSSTS ACTVSSKVAE KAEISLCGRS LTNPVSVRI MLHAGLMYSR  
 301 RAVVSRVAKS WSFAYMPDLV SRLNRLDLPT LV\*

Based on the putative transmembrane domain in the gonococcal protein, it is predicted that the  
 15 proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for  
 vaccines or diagnostics, or for raising antibodies.

### Example 18

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 147>:

1 ..GGCCATTACT CCGACCGCAC TTGGAAGCCG CGTTTGNCG GCCGCCGTCT  
 20 51 GCCGTATCTG CTTTATGGCA CGCTGATTGC GGTATTGTG ATGATTTTGA  
 101 TGCCGAATC GGCAGCTTC GGTTCGGCT ATGCGTCGCT GCGCGCTTTG  
 151 TCGTTCGGCG CGCTGATGAT TGCGCTGTTA GACGTGTCGT CAAATATGGC  
 201 GATGCAGCCG TTTAAGATGA TGGTCGGCGA CATGGTCAAC GAGGAGCAGA  
 251 AAA.NTACGC CTACGGGATT CAAAGTTTCT TAGCAAATAC GGGCGCGGTC  
 25 301 GTGGCGGCGA TTCTGCCGTT TGTGTTTGGC TATATCGGTT TGGCGAACAC  
 351 CGCCGANAAA GCGGTTGTGC CGCAGACCGT GGTCTGGCGC TTTTATGTGG  
 401 GTGCGGCGTT GCTGGTGATT ACCAGCGCGT TCACGATTTT CAAAGTGAAG  
 451 GAATACGANC CGGAAACCTA CGCCCGTTAC CACGGCATCG ATGTCGCCGC  
 501 GAATCAGGAA AAAGCCAACT GGATCGCACT CTAAAA.CC GCGC..

30 This corresponds to the amino acid sequence <SEQ ID 148; ORF16>:

1 ..GHYSDRTWKP RLXGRRLPYL LYGTLIAVIV MILMPNSGSF GFGYASLAAL  
 51 SFGALMIALL DVSSNMAMQP FKMMVGDVMN EEQKXYAYGI QSFLANTGAV  
 101 VAILPFVFA YIGLANTAXK GVVPQTVVVA FYVGAALLVI TSAFTIFKVK  
 151 EYXPETYARY HGIDVAANQE KANWIALLKX A..

35 Further work revealed the complete nucleotide sequence <SEQ ID 149>:

1 ATGTCGGAAT ATACGCCTCA AACAGCAAAA CAAGGTTTGC CCGCGCTGGC  
 51 AAAAAGCACG ATTTGGATGC TCAGTTTCGG CTTTCTCGGC GTTCAGACGG  
 101 CCTTTACCTT GCAAAGCTCG CAAATGAGCC GCATTTTTC AAGCTAGGC  
 151 GCAGACCCGC ACAATTTGGG CTGGTTTTTC ATCCTGCCGC CGCTGGCGGG  
 201 GATGCTGGTG CAGCCGATTG TCGGCCATTA CTCCGACCGC ACTTGAAGC  
 251 CGCGTTTGGG CCGCGCCCGT CTGCGGTATC TGCTTTATGG CACGCTGATT  
 301 GCGGTTATTG TGATGATTTT GATGCCGAAC TCGGCGAGCT TCGGTTTCGG  
 351 CTATGCGTCG CTGCGCGGCT TGTGTTTCGG CGCGCTGATG ATTGCGCTGT  
 401 TAGACGTGTC GTCAAATATG GCGATGCAGC CGTTTAAGAT GATGGTCGGC  
 451 GACATGGTCA ACGAGGAGCA GAAAGGCTAC GCCTACGGGA TTCAAAGTTT  
 501 CTTAGCAAT ACGGGCGCGG TCGTGCGCGC GATTCTGCCG TTTGTGTTTG  
 551 CGTATATCGG TTTGGCGAAC ACCGCCGAGA AAGGCGTTGT GCCGAGACC  
 601 GTGGTCGTGG CGTTTATGT GGGTCGGCGC TTGCTGGTGA TTACCAGCGC  
 651 GTTCACGATT TTCAAAGTGA AGGAATACGA TCCGGAACC TACGCCCGTT  
 701 ACCACGGCAT CGATGTCGCC GCGAATCAGG AAAAAGCCAA CTGGATCGAA  
 751 CTCTTGAAAA CCGCGCCTAA GCGGTTTTGG ACGGTTACTT TGGTGCAATT  
 801 CTTCTGCTGG TTCGCCCTCC AATATATGTG GACTTACTCG GCAGGCGCGA  
 851 TTGCGGAAAA CGTCTGGCAC ACCACCGATG CGTCTCCGT AGGTTATCAG  
 901 GAGCGGGTA ACTGGTACGG CGTTTGGCG GCGGTGCACT CCGTTGCGGC

5  
 951 GGTGATTTGT TCGTTTGTAT TGGCGAAAGT GCCGAATAAA TACCATAAGG  
 1001 CGGGTTATTT CGGCTGTTTG GCTTTGGGCG CGCTCGGCTT TTTCTCCGTT  
 1051 TTCTTCATCG GCAACCAATA CGCGCTGGTG TTGTCTTATA CCTTAATCGG  
 1101 CATCGCTTGG GCGGGCATT TCACTTATCC GCTGACGATT GTGACCAACG  
 1151 CCTTGTGCGG CAAGCATATG GGCACCTACT TGGGCTTGTT TAACGGCTCT  
 1201 ATCTGTATGC CTCAAATCGT CGCTTCGCTG TTGAGTTTCG TGCTTTTCCC  
 1251 TATGCTGGGC GGCTTGACAG CCACTATGTT CTTGGTAGGG GCGCTCGTCC  
 1301 TGCTGCTGGG CGCGTTTTC GTGTTCTCTGA TTAAAGAAAC ACACGGCGGG  
 1351 GTTTGA

10 This corresponds to the amino acid sequence <SEQ ID 150; ORF16-1>:

1 MSEYTPQTAK QGLPALAKST IWMLSFGFLG VQTAFTLQSS QMSRIFQTLG  
 51 ADPHNLGWFF ILPPLAGMLV QPIVGHYSDR TWKPRLGRR LPYLLYGTLI  
 101 AVIVMILMPN SGSFGFGYAS LAALSFGALM IALLDVSSNM AMQPFKMMVG  
 151 DMVNEEQKY AYGIQSFLAN TGAVVAAILP FVFAYIGLAN TAEKGVVPQT  
 201 VVVAFYVGAA LLVITSFTI FKVKEYDPET YARYHGIDVA ANQEKANWIE  
 251 LLKTAPKAFW TVTLVQFFCW FAFQYMWYTS AGAIAENVWH TTDASSVGQ  
 301 EAGNWWYVLA AVQSVAVIC SFVLAKVPNK YHKAGYFGCL ALGALGFFSV  
 351 FFIGNQYALV LSYTLIGIAW AGIITYPLTI VTNALSGKHM GTYLGFLNGS  
 401 ICMPIVASL LSFVLFPLMG GLQATMFLVG GVVLLLGAFS VFLIKETHGG  
 451 V\*

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF16 shows 96.7% identity over a 181aa overlap with an ORF (ORF16a) from strain A of *N. meningitidis*:

25  
 orf16.pep GHYSDRTWKPRXLGRRLLPYLLYGTLIIV  
 orf16a IFQTLGADPHSLGWFFILPPLAGMLVQPIVGHYSDRTWKPRLGRRLLPYLLYGTLIIV  
 30  
 orf16.pep MILMPNSGSFGFGYASLAALSFGALMIALLDVSSNMAMQPFKMMVGMVNEEQKYAYGI  
 orf16a MILMPNSGSFGFGYASLAALSFGALMIALLDVSSNMAMQPFKMMVGMVNEEQKYAYGI  
 35  
 orf16.pep QSFLANTGAVVAAILPFVFAYIGLANTAXKGVVPQTVVVAFYVGAALLVITSFTI FKVK  
 orf16a QSFLANTGAVVAAILPFVFAYIGLANTAEKGVVPQTVVVAFYVGAALLVITSFTI FKVK  
 40  
 orf16.pep EYXPETYARYHGIDVAANQEKANWIALKXA  
 orf16a EYNPETYARYHGIDVAANQEKANWIELLKTAPKAFWTVTLVQFFCWFAFQYMWYTSAGAI  
 45  
 orf16a AENVWHTTDASSVGQAEAGNWWYVLA AVQSVAVIC SFVLAKVPNK YHKAGYFGCLALGA  
 50

The complete length ORF16a nucleotide sequence <SEQ ID 151> is:

55  
 1 ATGTCGGAAT ATACGCCTCA AACAGCAAAA CAAGGTTTGC CCGCGCTGGC  
 51 AAAAAGCAGC ATTTGGATGC TCAGTTTCGG CTTTCTCGGC GTTCAGACGG  
 101 CCTTTACCCT GCAAAGCTCG CAGATGAGCC GCATCTTCCA GACGCTCGGT  
 151 GCCGATCCGC ACAGCCTCGG CTGGTTCTTT ATCCTGCCGC CGCTGGCGGG  
 201 GATGCTGGTG CAGCCGATTG TCGGCCATTA CTCGACCGC ACTTGAAGC  
 251 CGCGTTTGGG CGGCGCGCGT CTGCCGTATC TGCTTTATGG CACGCTGATT  
 301 GCGGTTATTG TGATGATTTT GATGCCGAAC TCGGCGAGCT TCGGTTTCGG  
 351 CTATGCGTCG CTGGCGGCTT TGTCGTTCCG CGCGCTGATG ATTGCGCTGT  
 60 401 TAGACGTGTC GTCAAATATG GCGATGCAGC CGTTAAGAT GATGGTCGGC

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65  
70  
75  
80  
85  
90  
95  
100  
105  
110  
115  
120  
125  
130  
135

451 GACATGGTCA ACGAGGAGCA GAAAGGCTAC GCCTACGGGA TTCAAAGTTT  
501 CTTAGCGAAT ACGGGCGCGG TCGTGGCGGC GATTCTGCCG TTTGTGTTTG  
551 CGTATATCGG TTTGGCGAAC ACCGCCGAGA AAGGCGTTGT GCCGCAGACC  
601 GTGGTCGTGG CGTTTTATGT GGGTGC GGCG TGTCTGGTGA TTACCAGCGC  
651 GTTCACGATT TTCAAAGTGA AGGAATACAA TCCGGAAACC TACGCCCGTT  
701 ACCACGGCAT CGATGTCGCC GCGAATCAGG AAAAAGCCAA CTGGATCGAA  
751 CTCTTGAAAA CCGCGCCTAA GCGGTTTGG ACGGTTACTT TGGTGCAATT  
801 CTTCTGCTGG TTCGCCTTCC AATATATGTG GACTTACTCG GCAGGCGCGA  
851 TTGCGGAAAA CGTCTGGCAC ACCACCGATG CGTCTTCCGT AGGTTATCAG  
901 GAGGCGGGTA ACTGGTACGG CGTTTTGGCG GCGGTGCAGT CGGTTGCGGC  
951 GGTGATTTGT TCGTTTGTAT TGGCGAAAGT GCCGAATAAA TACCATAAGG  
1001 CGGGTTATTT CGGCTGTTTG GCTTTGGGCG CGCTCGGCTT TTTCTCCGTT  
1051 TTCTTCATCG GCAACCAATA CGCGCTGGTG TTGTCTTATA CCTTAATCGG  
1101 CATCGCTTGG GCGGGCATTG TCACCTATCC GCTGACGATT GTGACCAACG  
1151 CCTGTGTCGG CAAGCATATG GGCACCTACT TGGGCCTGTT TAACGGCTCT  
1201 ATCTGTATGC CGCAAATCGT CGCTTCGCTG TTGAGTTTCG TGCTTTTCCC  
1251 TATGCTGGGC GGCTTGCAGG CCACTATGTT CTGGTAGGG GCGCTCGTCC  
1301 TGCTGCTGGG CGCGTTTTCG GTGTTCTCTGA TTAAAGAAAC ACACGGCGGG  
1351 GTTTGA

20 This encodes a protein having amino acid sequence <SEQ ID 152>:

1 MSEYTPQTAK QGLPALAKST IWMLSFGLG VQTAFTLQSS QMSRIFQTLG  
51 ADPHSLGWFF ILPLAGMLV QPIVGHYSR TWKPRLGRR LPYLLYGLLI  
101 AVIVMILMPN SGSFGFGYAS LAALSFGALM IALLDVSSNM AMQPFKMMVG  
151 DMVNEEQKY AYGIQSFLAN TGAVVAAILP FVFAYIGLAN TAEKGVPQT  
201 VVVAFYVGAA LLVITSFTI FKVKEYNPET YARYHGIDVA ANQEKANWIE  
251 LLKTAPKAFW TVTLVQFFCW FAFQYMWYTS AGAIAENVWH TTDASSVGYQ  
301 EAGNWyGVLA AVQSVAVIC SFVLAKVPNK YHKAGYFGCL ALGALGFFSV  
351 FFIGNQYALV LSYTLIGIAW AGIITYPLTI VTNALSGKHM GTYLGFLFNGS  
401 ICMPQIVASL LSFVLPMLG GLQATMFLVG GVVLLLGAFS VFLIKETHGG  
451 V\*

ORF16a and ORF16-1 show 99.6% identity in 451 aa overlap:

35  
40  
45  
50  
55  
60  
65

orf16a.pep	10	20	30	40	50	60
orf16-1	10	20	30	40	50	60
orf16a.pep	70	80	90	100	110	120
orf16-1	70	80	90	100	110	120
orf16a.pep	130	140	150	160	170	180
orf16-1	130	140	150	160	170	180
orf16a.pep	190	200	210	220	230	240
orf16-1	190	200	210	220	230	240
orf16a.pep	250	260	270	280	290	300
orf16-1	250	260	270	280	290	300
orf16a.pep	310	320	330	340	350	360
orf16-1	310	320	330	340	350	360

		370	380	390	400	410	420
	orf16a.pep	LSYTLIGIAWAGIITYPLTIVTNALSGKHMGTYLGLFNGSICMPQIVASLLSFVLPMLG					
5	orf16-1	LSYTLIGIAWAGIITYPLTIVTNALSGKHMGTYLGLFNGSICMPQIVASLLSFVLPMLG					
		370	380	390	400	410	420
		430	440	450			
10	orf16a.pep	GLQATMFLVGGVVLLLGAFSVFLIKETHGGVX					
	orf16-1	GLQATMFLVGGVVLLLGAFSVFLIKETHGGVX					
		430	440	450			

### Homology with a predicted ORF from *N.gonorrhoeae*

- 15 ORF16 shows 93.9% identity over a 181aa overlap with a predicted ORF (ORF16.ng) from *N. gonorrhoeae*:

	orf16.pep	GHYSDRTWKPRXLGRRLPYLLYGTLIAVIV	30
	orf16ng	HFSNARRRPAQFGLVFHPAAAGGDAGSADSGYSDRTWKPRLGRRLPYLLYGTLIAVIV	131
20	orf16.pep	MILMPNSGSFSGFYASLAALSFGALMIALLDVSSNMAMQPFKMMVGMVNEEQKXYAYGI	90
	orf16ng	MILMPNSGSFSGFYASLAALSFGALMIALLDVSSNMAMQPFKMMVGMVNEEQKSYAYGI	191
25	orf16.pep	QSFLANTGAVVAAILPFVFAYIGLANTAXKGVVPQTVVVAFYVGAALLVITSAFTIFKVK	150
	orf16ng	QSFLANTDAVVAAILPFVFAYIGLANTAEGKGVVPQTVVVAFYVGAALLIITSAFTISKVK	251
30	orf16.pep	EYXPETYARYHGIDVAANQEKANWIALLKXA	181
	orf16ng	EYDPETYARYHGIDVAANQEKANWFELLKTAPKVFWTVPVQFFCWFAFRYMWTYSAGAI	311

The complete length ORF16ng nucleotide sequence <SEQ ID 153> is:

	1	ATGATAGGGG	ATCGCCGCGC	CGGCAACCAT	TTCGGATTTT	CCAAAGCAAA
	51	TACTTTTCAA	ATCAAAAAAA	AGGATTTACT	TTATGTCGGA	ATATACGCCT
35	101	CAACACAGCA	AACAAGGTTT	GCCCGCGCCG	GCAAAAAGCA	CGATTGGGAT
	151	GTTGAGCTTC	GGCTATCTCG	GCGTTCAGAC	GGCCTTTACC	CTGCAAAGCT
	201	CGCAGATGAG	CCGCATTTT	CAAACGCTAG	GCGCAGACCC	GCACAATTG
	251	GGCTGGTTTT	TCATCCTGCC	GCCGCTGGCG	GGGATGCTGG	TTACGCCGAT
40	301	AGTGGCTACT	ACTCAGACCG	CACTTGGAAG	CCGCGCTTGG	GCGGCCGCCG
	351	CCTGCCGTAT	CTGCTTTACG	GCACGCTGAT	TGCGGTCATC	GTGATGATTT
	401	TGATGCCGAA	CTCGGGCAGC	TTCGGTTTCG	GCTATGCGTC	GCTGGCGGCC
	451	TTGTGCTTCG	GCGCGCTGAT	GATTGCGCTG	TTGGACGTGT	CGTCGAATAT
	501	GGCGATGCAG	CCGTTTAAGA	TGATGGTCGG	CGATATGGTC	AACGAGGAGC
45	551	AGAAAAGCTA	CGCTACGGG	ATTCAAAGTT	TCTTAGCGAA	TACGGACGCG
	601	GTTGTGGCAG	CGATTCTGCC	GTTTGTGTTT	GCGTATATCG	GTTTGGCGAA
	651	CACTGCCGAG	AAAGGCGTTG	TGCCACAAAC	CGTGGTCGTA	GCATTCTATG
	701	TGGGTGCGGC	GTTACTGATT	ATTACCAAGT	CGTTCACAAT	CTCCAAAGTC
	751	AAAGAATACG	ACCCGGAAC	CTACGCCCGT	TACCACGGCA	TCGATGTCGC
	801	CGCGAATCAG	GAAAAAGCCA	ACTGGTTCGA	ACTCTTAAAA	ACCGCGCCTA
50	851	AAGTGTTTTG	GACGGTTACT	CCGGTACAGT	TTTTCTGCTG	GTTTCGCCTTC
	901	CGGTATATGT	GGACTTACTC	GGCAGGCGCG	ATTGCAGAAA	ACGTCTGGCA
	951	CACTACCGAT	GCGTCTCCG	TAGGCCATCA	GGAGCGGGC	AACCGGTACG
	1001	GCGTTTTGGC	GGCGGTGTAG			

This encodes a protein having amino acid sequence <SEQ ID 154>:

55	1	MIGDRRAGNH	FGFSKANTFQ	IKKKDLLYVG	IYASNSKTRF	ARAGKKHDL
	51	VELRLSRRSD	GLYPAKLAD	PHFSNARRRP	AQFGLVFHFA	AAGGDAGSAD
	101	SGYSDRTWK	PRLGRRRLPY	LLYGTLIAVI	VMILMPNSGS	FGFGYASLAA
	151	LSFGALMIAL	LDVSSNMAMQ	PFKMMVGMV	NEEQKSYAYG	IQSFLANTDA
60	201	VVAAILPFVF	AYIGLANTAE	KGVVPQTVV	AFYVGAALLI	ITSFTISKV
	251	KEYDPETYAR	YHGIDVAANQ	EKANWFELLK	TAPKVFWTVT	PVQFFCWFAF
	301	RYMWTYSAGA	IAENVHHTD	ASSVGHQEAG	NRYGVLAAV*	

ORF16ng and ORF16-1 show 89.3% identity in 261 aa overlap:

		30	40	50	60	70	80
	orf16-1.pep	MLSFGFLGVQTAFTLQSSQMSRIFQTLGADPHNLGWFFILPPLAGMLVQPI-VGHYSDRT					
5	orf16ng	DVELRLSRSDGLYPAKLADEPHFSNARRRPAQFGLVF-HPAAAGGDAGSADSGYYSVRT					
		50	60	70	80	90	100
		90	100	110	120	130	140
10	orf16-1.pep	WKPRLGGRRLPYLLYGTLIAVIVMILMPNSGSFGFGYASLAALSFGALMIALLDVSSNMA					
	orf16ng	WKPRLGGRRLPYLLYGTLIAVIVMILMPNSGSFGFGYASLAALSFGALMIALLDVSSNMA					
		110	120	130	140	150	160
		150	160	170	180	190	200
15	orf16-1.pep	MQPFKMMVGDVNEEQKGYAYGIQSFLANTGAVVAAILPFVFAYIGLANTAEGVVPQTV					
	orf16ng	MQPFKMMVGDVNEEQKSYAYGIQSFLANTDAVVAAILPFVFAYIGLANTAEGVVPQTV					
		170	180	190	200	210	220
		210	220	230	240	250	260
20	orf16-1.pep	VVAFYVGAALLVITSFTIFKVKEYDPEYARYHGIDVAANQEKANWIELLKTAAPKFWT					
	orf16ng	VVAFYVGAALLIITSFTISKVKEYDPEYARYHGIDVAANQEKANWFELLKTAAPKFWT					
		230	240	250	260	270	280
		270	280	290	300	310	320
	orf16-1.pep	VTLVQFFCWFQFYMWYTSAGAIENVWHTDASSVGYQEAGNRYGVLAQVAAVICS					
	orf16ng	VTPVQFFCWFQFRYMWYTSAGAIENVWHTDASSVGHQEAGNRYGVLAQVAVX					
30		290	300	310	320	330	340

Based on this analysis, including the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### 35 Example 19

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 155>:

	1	ATGTTGTTCC	GTAAACGAC	CGCCGCCGTT	TTGGCGCATA	CCTTGATGCT
	51	GAACGGCTGT	ACGTTGATGT	TGTGGGGAAT	GAACAACCCG	GTCAGCGAAA
40	101	CAATCACCCG	NAAACACGTT	GNCAAAGACC	AAATCCGNGN	CTTCGGTGTG
	151	GTTGCCGAAG	ACAATGCCCA	ATTGGAAAAG	GGCAGCCTGG	TGATGATGGG
	201	CGGAAAATAC	TGGTTCGTCG	TCAATCCCGA	AGATTCCGCG	AA.NTGACGG
	251	GNATTTTGAN	GGCAGGGCTG	GACAAACCCT	TCCAAATAGT	TNAGGATACC
	301	CCGAGCTATG	C.TGCCACCA	AGCCCTGCCG	GTCAAACTCG	GATCGNCTGG
	351	CAGCCAGAAT...				

45 This corresponds to the amino acid sequence <SEQ ID 156; ORF28>:

	1	MLFRKTTAAV	LAHTLMLNGC	TLMLWGMNPN	VSETITRKHV	XKDQIRXFGV
	51	VAEDNAQLEK	GSLVMMGGKY	WFVVPEDSA	XTGILXAGL	DKPFQIVXDT
	101	PSYXCHQALP	VKLGSXGSQN...			

Further work revealed the complete nucleotide sequence <SEQ ID 157>:

50	1	ATGTTGTTCC	GTAAACGAC	CGCCGCCGTT	TTGGCGGCAA	CCTTGATGCT
	51	GAACGGCTGT	ACGTTGATGT	TGTGGGGAAT	GAACAACCCG	GTCAGCGAAA
	101	CAATCACCCG	CAAACACGTT	GACAAAGACC	AAATCCGCGC	CTTCGGTGTG
	151	GTTGCCGAAG	ACAATGCCCA	ATTGGAAAAG	GGCAGCCTGG	TGATGATGGG
	201	CGGAAAATAC	TGGTTCGTCG	TCAATCCCGA	AGATTCCGCG	AAGCTGACGG
55	251	GCATTTTGAA	GGCAGGGCTG	GACAAACCCT	TCCAAATAGT	TGAGGATACC

5  
301 CCGAGCTATG CTCGCCACCA AGCCCTGCCG GTCAAACCTCG AATCGCCTGG  
351 CAGCCAGAAT TTCAGTACCG AAGGCCTTTG CCTGCGCTAC GATACCGACA  
401 AGCCTGCCGA CATCGCCAAG CTGAAACAGC TCGGGTTTGA AGCGGTCAAA  
451 CTCGACAATC GGACCATTTA CACGCGCTGC GTATCCGCCA AAGGCAAATA  
501 CTACGCCACA CCGCAAAAAC TGAACGCCGA TTACCATTTT GAGCAAAGTG  
551 TGCCTGCCGA TATTTATTAC ACGGTTACTG AAGAACATAC CGACAAATCC  
601 AAGCTGTTTG CAAATATCTT ATATACGCCC CCCTTTTGA TACTGGATGC  
651 GCGGGGCGCG GTACTGGCCT TGCCTGCGGC GGCTCTGGGT GCGGTCGTGG  
701 ATGCCGCCCG CAAATGA

10 This corresponds to the amino acid sequence <SEQ ID 158; ORF28-1>:

15  
1 MLFRKTTAAV LAATLMLNGC TMLWGMNPN VSETITRKHV DKDQIRAFGV  
51 VAEDNAQLEK GSLVMMGGKY WVVNPEDSA KLTGILKAGL DKPFQIVEDT  
101 PSYARHQALP VKLESPGSQN FSTEGCLLRY DTDKPADIAP LKQLGFEAVK  
151 LDNRTIYTRC VSAKGKYYAT PQKLNADYHF EQSVPADIYY TVTEEHTDKS  
201 KLFANILYTP PFLILDAAGA VLALPAAALG AVVDAARK\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF28 shows 79.2% identity over a 120aa overlap with an ORF (ORF28a) from strain A of *N. meningitidis*:

20  
or28.pep 10 20 30 40 50 60  
MLFRKTTAAVLAHTLMLNGCTMLWGMNPNVSETITRKHVXKDQIRXFGVVAEDNAQLEK  
or28a 10 20 30 40 50 60  
MLFRKTTAAVLAATLMLNGCTVMMWGMNSPFSETTARKHVDKDQIRAFGVVAEDNAQLEK  
25  
or28.pep 70 80 90 100 110 120  
GSLVMMGGKYWVVNPEDSAXXTGILXAGLDKPFQIVXDTPSYXCHQALPVKLGSGSQN  
or28a 70 80 90 100 110 120  
GSLVMMGGKYWVVNPEDSAKLTGILKAGLDKQFQMVPEPNPFA-YQALPVKLESPASQN  
30  
or28a 120 130 140 150 160 170  
FSTEGCLLRYDTRDPADIAPKLQLEFEAVELDNRTIYTRCVSAKGKYYATPQKLNADYHF

The complete length ORF28a nucleotide sequence <SEQ ID 159> is:

35  
1 ATGTTGTTCC GTAAAACGAC CGCCGCCGTT TTGGCGGCAA CTTGATGTT  
51 GAACGGCTGT ACGGTAATGA TGTGGGGTAT GAACAGCCCG TTCAGCGAAA  
101 CGACCGCCCG CAAACACGTT GACAAGGACC AAATCCGCGC CTCGCGTGTG  
151 GTTGCCGAAG ACAATGCCCA ATTGGAAGG GGCAGCCTGG TGATGATGGG  
201 CGGGAAATAC TGGTTCGTCG TCAATCCTGA AGATTCCGCG AAGCTGACGG  
40  
251 GCATTTTGAA GGCCGGGTG GACAAGCAGT TTCAAATGGT TGAGCCCAAC  
301 CCGCGCTTTG CCTACCAAGC CTGCGCGGTC AACTCGAAT CGCCCGCCAG  
351 CCAGAATTC AGTACCGAAG GCCTTGCCT GCGCTACGAT ACCGACAGAC  
401 CTGCCGACAT CGCCAAGCTG AACAGCTTG AGTTGAAGC GGTGCAACTC  
451 GACAATCGGA CCATTTACAC GCGCTGCGTC TCCGCCAAAG GCAAACTACTA  
45  
501 CGCCACACCG CAAAACCTGA ACGCCGATTA TCATTTTGAG CAAAGTGTGC  
551 CTGCCGATAT TTATTACAC GTTACGAAA AACATACCGA CAAATCCAAG  
601 TTGTTTGAAA ATATTGCATA TACGCCACC ACGTTGATAC TGGATGCGGT  
651 GGGCGCGGTG CTGGCCTTGC CTGTCGCGGC GTTGATTGCA GCCACGAATT  
701 CCTCAGACAA ATGA

50 This encodes a protein having amino acid sequence <SEQ ID 160>:

55  
1 MLFRKTTAAV LAATLMLNGC TVMMWGMNSP FSETTARKHV DKDQIRAFGV  
51 VAEDNAQLEK GSLVMMGGKY WVVNPEDSA KLTGILKAGL DKQFQMVPEPN  
101 PRFAYQALPV KLESPASQNF STEGLCLRYD TDRPADIAPL KQLEFEAVEL  
151 LDNRTIYTRCV SAKGKYYATP QKLNADYHFE QSVPAIYYT VTKKHTDKSK  
201 LFENIAYTPT TLILDVAVGAL LALPVAALIA ATNSSDK\*

ORF28a and ORF28-1 show 86.1% identity in 238 aa overlap:



		10	20	30	40	50	60
	orf28a.pep	MLFRKTTAAVLAATLMLNGCTVMMWGMNSPFSETTARKHVDKDQIRAFGVVAEDNAQLEK					
5	orf28-1	MLFRKTTAAVLAATLMLNGCTLMLWGMNPNVSETITRKHVDKDQIRAFGVVAEDNAQLEK					
		10	20	30	40	50	60
	orf28a.pep	70	80	90	100	110	119
10	orf28-1	GSLVMMGGKYWFVNPEDSAKLTGILKAGLDKQFQMVFNPRFA-YQALPVKLESPASQN					
		70	80	90	100	110	120
	orf28a.pep	120	130	140	150	160	179
15	orf28-1	FSTEGLCRLRYDTRPADIAKLKQLEFEAVELDNRTIYTRCVSAKGKYYATPQKLNADYHF					
		130	140	150	160	170	180
	orf28a.pep	180	190	200	210	220	230
20	orf28-1	EQSVPADIIYTVTKKHTDKSKLFENIAYTPTTLILDVAGAVLALPVAALIAATNSSDKX					
		190	200	210	220	230	

### 25 Homology with a predicted ORF from *N.gonorrhoeae*

ORF28 shows 84.2% identity over a 120aa overlap with a predicted ORF (ORF28.ng) from *N.gonorrhoeae*:

	orf28.pep	MLFRKTTAAVLAHTLMLNGCTLMLWGMNPNVSETITRKHVXKDQIRXFGVVAEDNAQLEK	60
30	orf28ng	MLFRKTTAAVLAATLILNGCTMMLRGMNPNVSETITRKHVDKDQIRAFGVVAEDNAQLEK	60
	orf28.pep	GSLVMMGGKYWFVNPEDSAXXTGILXAGLDKPFQIVXDTPSYXCHQALPVKLGSGXGSON	120
	orf28ng	GSLVMMGGKYWFAVNPEDSAKLTGLLKAGLDKPFQIVEDTPSYARHQALPVKFEAPGSON	120

### 35 The complete length ORF28ng nucleotide sequence <SEQ ID 161> is

	1	ATGTTGTTCC	GTAAACGAC	CGCCGCCGTT	TTGGCGGCAA	CCTTGATACT
	51	GAACGGCTGT	ACGATGATGT	TGCGGGGGAT	GAACAACCCG	GTCAGCCAAA
	101	CAATCACCCG	CAACACGTT	GACAAAGACC	AAATCCGCGC	CTTCGGTGTG
40	151	GTTGCCGAAG	ACAATGCCCA	ATTGGAAGAG	GGCAGCCTGG	TGATGATGGG
	201	CGGGAAATAC	TGGTTCGCGG	TCAATCCCGA	AGATTCCGCG	AAGCTGACGG
	251	GCCTTTTGAA	GGCCGGGTTG	GACAAGCCCT	TCCAAATAGT	TGAGGATACC
	301	CCGAGCTATG	CCCGCCACCA	AGCCCTGCCG	GTCAAATTCG	AAGCGCCCGG
	351	CAGCCAGAAT	TTCAGTACCG	GAGGTCTTTG	CCTGCGCTAT	GATACCGGCA
45	401	GACCTGACGA	CATCGCCAAG	CTGAAACAGC	TTGAGTTTAA	AGCGGTCAAA
	451	CTCGACAATC	GGACCATTTA	CACGCGCTGC	GTATCCGCCA	AAGGCAATA
	501	CTACGCCACG	CCGCAAAAAC	TGAACGCCGA	TTATCATTTT	GAGCAAAGTG
	551	TGCCC GCCGA	TATTTATTAT	ACGGTTACTG	AAAAACATAC	CGACAAATCC
	601	AAGCTGTTTG	GAATATCTT	ATATACGCCC	CCCTTGTTGA	TATTGGATGC
50	651	GGCGGCCGCG	GTGCTGGTCT	TGCCTATGGC	TCTGATTGCA	GCCGCGAATT
	701	CCTCAGACAA	ATGA			

This encodes a protein having amino acid sequence <SEQ ID 162>:

	1	MLFRKTTAAV	LAATLILNGC	TMMLRGMNPN	VSQTITRKHV	DKDQIRAFGV
	51	VAEDNAQLEK	GSLVMMGGKY	WFAVNPEDSA	KLTGLLKAGL	DKPFQIVEDT
55	101	PSYARHQALP	VKFEAPGSON	FSTGGLCLRY	DTGRPDIAK	LKQLEFKAVK
	151	LDNRTIYTRC	VSAKGKYYAT	PQKLNADYHF	EQSVPADIIY	TVTEKHTDKS
	201	KLFGNILYTP	PLILDAAAA	VLVLPALIA	AANSSDK*	

ORF28ng and ORF28-1 share 90.0% identity in 231 aa overlap:

		10	20	30	40	50	60
60	orf28-1.pep	MLFRKTTAAVLAATLMLNGCTLMLWGMNPNVSETITRKHVDKDQIRAFGVVAEDNAQLEK					
		:     :     :     :     :     :					

	orf28ng	MLFRKTTAAVLAATLILNGCTMMLRGMNPNVSQTITRKHVVDKQIRAFGVVAEDNAQLEK	10	20	30	40	50	60
5	orf28-1.pep	GSLVMMGGKYWFVNPEDSAKLTGILKAGLDKPFQIVEDTPSYARHQALPVKLESPGSQN	70	80	90	100	110	120
	orf28ng	GSLVMMGGKYWFAVNPEDSAKLTGLLKAGLDKPFQIVEDTPSYARHQALPVKFEAPGSQN	70	80	90	100	110	120
10	orf28-1.pep	FSTEGLCRLRYDTDKPADIAKLKQLGFEAVKLDNRTIYTRCVSAKGKYYATPQKLNADYHF	130	140	150	160	170	180
	orf28ng	FSTGGLCLRYDTGRPDDIAKLKQLEFKAVKLDNRTIYTRCVSAKGKYYATPQKLNADYHF	130	140	150	160	170	180
15	orf28-1.pep	EQSVPADIIYTVTEEHTDKSKLFANILYTPPFLILDAAGAVLALPAAALGAVVDAARKX	190	200	210	220	230	239
20	orf28ng	EQSVPADIIYTVTEKHTDKSKLFGNILYTPPLLILDAAGAVLVLPMALIAAANSDDKX	190	200	210	220	230	

Based on this analysis, including the presence of a putative transmembrane domain in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

- 25 ORF28-1 (24kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 6A shows the results of affinity purification of the GST-fusion protein, and Figure 6B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA, which gave a positive result. These experiments confirm
- 30 that ORF28-1 is a surface-exposed protein, and that it may be a useful immunogen.

### Example 20

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 163>:

	1	..GTCAGTCCTG TACTGCCTAT TACACACGAA CGGACAGGGT TTGAAGGTGT
35	51	TATCGGTTAT GAAACCCATT TTTCAGGGCA CGGACATGAA GTACACAGTC
	101	CGTTCGATCA TCATGATTCA AAAAGCACTT CTGATTTTCAG CGGCGGTGTA
	151	GACGGCGGTT TTAAGTTTCA CCAACTTCAT CGAACATGGT CGGAAATCCA
	201	TCCGGAGGAT GAATATGACG GGCCGCAAGC AGCG.ATTAT CCGCCCCCG
	251	GAGGAGCAAG GGATATATAC AGCTATTATG TCAAAGGAAC TTCAACAAAA
40	301	ACAAAGACTA GTATTGTCCC TCAAGCCCCA TTTTCAGACC GTTGCTAGA
	351	AGAAAATGCC GGTGCCGCCT CTGGT..

This corresponds to the amino acid sequence <SEQ ID 164; ORF29>:

	1	..VSPVLPI THE RTGFEVGIVY ETHFSGHGHE VHSFPDHHDS KSTSDFSGGV
	51	DGGFTVYQLH RTWSEIHPED EYDGPQAAXY PPPGGARDIY SYYVKGTSTK
45	101	TKTSIVPQAP FSDRWLEENA GAASG..

- Further work revealed the complete nucleotide sequence <SEQ ID 165>:

	1	ATGAATTTGC CTATTCAAAA ATTCATGATG CTGTTTGCGAG CAGCAATATC
	51	GTTGCTGCAA ATCCCCATTA GTCATGCGAA CGGTTTGAT GCCCGTTTGC
	101	GCGATGATAT GCAGGCAGAAA CACTACGAAC CGGGTGGTAA ATACCATCTG
	151	TTTGGAATG CTCGCGGCAG TGTTAAAAAG CGGGTTTACG CCGTCCAGAC

5 201 ATTTGATGCA ACTGCGGTCA GTCCTGTACT GCCTATTACA CACGAACGGA  
 251 CAGGGTTTGA AGGTGTTATC GGTTATGAAA CCCATTTTTC AGGGCACGGA  
 301 CATGAAGTAC ACAGTCCGTT CGATCATCAT GATTCAAAAA GCACTTCTGA  
 351 TTTTCAGCGGC GGTGTAGACG GCGGTTTAC TGTTTACCAA CTTTCATCGAA  
 401 CAGGGTCGGA AATCCATCCG GAGGATGGAT ATGACGGGCC GCAAGGCAGC  
 451 GATTATCCGC CCCCCGAGG AGCAAGGGAT ATATACAGCT ATTATGTCAA  
 501 AGGAACTTCA AAAAAACAA AGACTAATAT TGTCCCTCAA GCCCCATTTT  
 551 CAGACCGTTG GCTAAAAGAA AATGCCGGTG CCGCCTCTGG TTTTTCAGC  
 601 CGTGCGGATG AAGCAGGAAA ACTGATATGG GAAAGCGACC CCAATAAAAA  
 10 651 TTGTGGGGCT AACCGTATGG ATGATGTTCC CGGCATCGTC CAAGGTGCGG  
 701 TTAATCCTTT TTAATGGGT TTTCAAGGAG TAGGGATTGG GGCAATTACA  
 751 GACAGTGCAG TAAGCCCCGT CACAGATACA GCCGCGCAGC AGACTCTACA  
 801 AGGTATTAAT GATTTAGGAA AATTAAGTCC GGAAGCACAA CTTGCTGCCG  
 851 CGAGCCTATT ACAGGACAGT GCTTTTCCGG TAAAAGACGG TATCAACTCT  
 15 901 GCCAAACAAT GGGCTGATGC CCATCCAAT ATAACAGCTA CTGCCCAAAC  
 951 TGCCCTTTCC GCAGCAGAGG CCGCAGGTAC GGTTTGGAGA GGTAATAAAG  
 1001 TAGAACTTAA CCGACTAAA TGGGATTGGG TAAAAATAC CGGTTATAAA  
 1051 AAACCTGCTG CCCGCCATAT GCAGACTTTA GATGGGGAGA TGGCAGGTGG  
 1101 GAATAAACCT ATTAATCTT TACCAAACAG TGCCGCTGAA AAAAGAAAAC  
 20 1151 AAAATTTTGA GAAGTTTAA AGTAAGTGA GTTCAGCAAG TTTTGATTCA  
 1201 GTGCACAAA CACTAACTCC CAATGCACCT GGTATTTTAA GTCCTGATAA  
 1251 AGTTAAACT CGATACACTA GTTAGATGG AAAAATTACA ATTATAAAG  
 1301 ATAACGAAA CAACTATTTT AGAATCCATG ATAATTACG AAAACAGTAT  
 25 1351 CTTGATTCAA ATGGTAATGC TGTGAAAACC GGTAATTAC AAGGTAAGCA  
 1401 AGCAAAGAT TATTACAAC AACAACTCA TATCAGGAAC TTAGACAAAT  
 1451 GA

This corresponds to the amino acid sequence <SEQ ID 166; ORF29-1>:

1 MNLPIQKFM LFAAISLLQ IPISHANGLD ARLRDDMQAK HYEPGGKYHL  
 51 FGNARGSVKK RVYAVQTFDA TAVSPVLPIT HERTGFEGVI GYETHFSGHG  
 101 HEVHSPFDHH DSKSTSDFSG GVDGGFTVYQ LHRTGSEIHP EDGYDGPQGS  
 151 DYPPPGGARD IYSYVKGTS TKTKTNIVPQ APFSDRWLKE NAGAASGFFS  
 201 RADEAGKLIW ESDPNKNWWA NRMDDVRGIV QGAVNPFLMG FQGVGIGAIT  
 251 DSAVSPVTDI AAQQTLOGIN DLGKLSPEAQ LAAASLLQDS AFAVKDGIN  
 301 AKQWADAHNP ITATAQTALS AEAAGTVWR GKVELNPTK WDVKNTGYK  
 351 KPAARHMOTL DGE MAGGNKP IKSLPNSAAE KRKQNFKEFN SNWSSAFDS  
 401 VHKTLPNAP GILSPDKVKT RYTSLDGKIT IIKDNENNYF RIHDNSRKQY  
 451 LDSNGNAVKT GNLQGKQAKD YLQQQTHIRN LDK\*

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

40 ORF29 shows 88.0% identity over a 125aa overlap with an ORF (ORF29a) from strain A of *N. meningitidis*:

	orf29.pep			10	20	30
				VSPVLPITHERTGFEGVIGYETHFSGHGHE		
45	orf29a	EPGGKYHLFGNARGSVKNRVYAVQTFDATAVGPILPITHERTGFEGIIGYETHFSGHGHE				
		50 60 70 80 90 100				
	orf29.pep		40 50 60 70 80 90			
50	orf29a	VHSPFDHHDHDSKSTSDFSGGVDGGFTVYQLHRTWSEIHPEDGYDGPQAXYPPPGGARDIY				
		110 120 130 140 150 160				
	orf29.pep		100 110 120			
55	orf29a	SYVVKGTSTKTKTSIVPQAPFSDRWLEENAGAASG				
		170 180 190 200 210 220				
60	orf29a	MDDIRGIVQGAVNPFLMGFQGVGIGAITDSAVSPVTDIAAQQTLOGXNHLGXLSPEAQLA				
		230 240 250 260 270 280				

The complete length ORF29a nucleotide sequence <SEQ ID 167> is:

5  
10  
15  
20  
25  
30

```

1  ATGAATTNGC CTATTCAAAA ATTCATGATG CTGTTTGCGAG CAGCAATATC
51  GTNGCTGCAA ATCCCNATTA GTCATGCGAA CGGTTTGGAT GCCCGTTTGC
101 GCGATGATAT GCAGGCAAAA CACTACGAAC CGGGTGGTAA ATACCATCTG
151 TTTGTAATG  CTCGCGGCAG TGTAAAAAAT CGGGTTTACG CCGTCCAAAC
201 ATTTGATGCA ACTGCGGTCG GCCCCATACT GCCTATTACA CACGAACGGA
251 CAGGATTTGA AGGCATTATC GGTATGAAA CCCATTTTTC AGGACATGGA
301 CATGAAGTAC ACAGTCCGTT CGATAATCAT GATTCAAAAA GCACTTCTGA
351 TTTCAGCGGC GCGTAGACG GTGGTTTTAC CGTTTACCAA CTTTCATCGGA
401 CAGGGTCGGA AATCCATCCG GAGGATGGAT ATGACGGGCC GCAAGGCAGC
451 GATTATCCGC CCCCCGGAGG AGCAAGGGAT ATATACANNT ANTATGTCAA
501 AGGAACTTCA ACAAAAAACA AGAGTAATAT TGTTCCCCGA GCCCATTTT
551 CAGACCGCTG GCTAAAAGAA AATGCCGGTG CCGCCTCTGG TTTTTCAGC
601 CGTGCTGATG AAGCAGGAAA ACTGATATGG GAAAGCGACC CCAATAAAAA
651 TTGGTGGGCT AACCGTATGG ATGATATTCG CGGCATCGTC CAAGTGCGG
701 TTAATCCTTT TTTAATGGGT TTCAAGGAG TAGGGATTGG GGCAATTACA
751 GACAGTGCAG TAAGCCCGGT CACAGATACA GCCGCGCAGC AGACTCTACA
801 AGGTATNAAT CATTTAGGAA ANTTAAGTCC CGAAGCACAA CTTGCGGCTG
851 CAACCGCATT ACAAGACAGT GCTTTTGC GG TAAAAGACGG TATCAATTCC
901 GCCAGACAAT GGGCTGATGC CCATCCGAAT ATAAGTCAA CAGCCCAAAC
951 TGCCCTTGCC GTAGCAGANG CCGCAACTAC GGTTTGGGGC GGTAATAAAG
1001 TAGAACTTAA CCCGACCAAA TGGGATTGGG TAAAAAATAC NGGCTATAAN
1051 ACACCTGCTG TTGCGACCAT GCATACTTGG GATGGGGAAA TGGCCGGTGG
1101 GAATAGACCG CCTAAATCTA TAACGTCCAA CAGCAAAGCA GATGCTTCCA
1151 CACAACCGTC TTTACAAGCG CACTAATTG GAGAACAAAT TANNNNNGGG
1201 CATGCTTATA ACAAGCATGT CATAAGACAA CAAGAATTTA CGGATTTAAA
1251 TATCAATTCA CCAGCAGATT TTGCTCGGCA TATTGAAAT ATTGTTAGCC
1301 ATCCANCAAA TATGAAAGAG TTACCTCGCG GTAGAACTGC GTATTGGGAT
1351 NATAAAACAG GGACNATAGT TATCCGAGAT AAAAATTCTG ACGATGGAGG
1401 TACAGCATT AGACCAACAT CAGGTAAAAA ATATTATGAT GATTTATAG

```

This encodes a protein having amino acid sequence <SEQ ID 168>:

35  
40

```

1  MNXPIQKFM LFAAAISXIQ IPISHANGLD ARLRDDMQAK HYEPPGKGYHL
51  FGNARGSVKN RVYAVQTFDA TAVGPILPIT HERTGFEGII GYETHFSGHG
101 HEVHSPFDNH DSKSTSDFSG GVDGGFTVYQ LHRTGSEIHP EDGYDGPQGS
151 DYPPPGGARD IYXXYVKGTS TKTKSNIVPR APFSDRWLKE NAGAASGFFS
201 RADEAGKLIW ESDPNKNWWA NRMDDIRGIV QGAVNPFLMG FQGVGIGAIT
251 DSAVSPVTD TAAQQLQGXN HLGXLSPEAQ LAAATALQDS AFAVKDGIN
301 ARQWADAHFN ITATAQTALA VAXAATTVWG GKVELNPTK WDWVKNTRYX
351 TPAVRTMHTL DGEMAGGNRP PKSITSNSKA DASTQPSLQA QLIGEIQIXG
401 HAYNKHVIRQ QEFTDLNINS PADEFARHIEN IVSHPXNMKE LPRGRYAYWD
451 KXTGTIVIRD KNSDDGGTAF RPTSGKYYD DL*

```

ORF29a and ORF29-1 show 90.1% identity in 385 aa overlap:

45  
50  
55  
60

```

              10      20      30      40      50      60
orf29a.pep  MNXPIQKFMMLFAAAISXLIQIPISHANGLDARLRDDMQAKHYEPGGKYHLFGNARGSVKN
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf29-1      MNLPIQKFMMLFAAAISLLQIPISHANGLDARLRDDMQAKHYEPGGKYHLFGNARGSVKK
              10      20      30      40      50      60

              70      80      90      100     110     120
orf29a.pep  RVYAVQTFDATAVGPILPITHERTGFEGIIIGYETHFSGHGHEVHSPFDNHDSKSTSDFSG
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf29-1      RVYAVQTFDATAVSPVLPIHERTGFEGVIGYETHFSGHGHEVHSPFDHDSKSTSDFSG
              70      80      90      100     110     120

              130     140     150     160     170     180
orf29a.pep  GVDGGFTVYQLHRTGSEIHPEDGYDGPQGS DYPPPGGARDIYXXYVKGSTKTKSNIVPR
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf29-1      GVDGGFTVYQLHRTGSEIHPEDGYDGPQGS DYPPPGGARDIYSYYVKGSTKTKTNIVPQ
              130     140     150     160     170     180

              190     200     210     220     230     240
orf29a.pep  APFSDRWLKENAGAASGFFSRADEAGKLIWESDPNKNWWANRMDDIRGIVQGAVNPFLMG
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf29-1      APFSDRWLKENAGAASGFFSRADEAGKLIWESDPNKNWWANRMDDVRGIVQGAVNPFLMG

```

		190	200	210	220	230	240
		250	260	270	280	290	300
5	orf29a.pep	FQGVGIGAITDS	AVSPVTD	TAAQQT	LQGXNHLG	XLSP	EALQDS
	orf29-1	FQGVGIGAITDS	AVSPVTD	TAAQQT	LQGXNHLG	XLSP	EALQDS
		250	260	270	280	290	300
10	orf29a.pep	ARQWADAH	PNITATA	QALAVAXA	ATTVWGGK	VELNPTK	WDWVKN
	orf29-1	AKQWADAH	PNITATA	QALAVAXA	ATTVWGGK	VELNPTK	WDWVKN
		310	320	330	340	350	360
15	orf29a.pep	DGEMAGGN	RPPK	SITSNS	KADAST	QPSLQA	LIGEXX
	orf29-1	DGEMAGGN	KPIKSLP	-NSAAE	KRKQNF	EKNSW	SSAS
		370	380	390	400	410	420

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF29 shows 88.8% identity over a 125aa overlap with a predicted ORF (ORF29.ng) from *N.gonorrhoeae*:

25	orf29.pep	VSPVLP	IHER	TGFEG	VIGY	ETHF	SGHGHE	30
	orf29ng	EPGGKYH	LFGN	ARGSV	KNRVC	AVQTF	DATA	102
30	orf29.pep	VHSPFD	HHD	SKST	SDFSG	VDGG	FTVY	90
	orf29ng	VHSPFD	HHD	SKST	SDFSG	VDGG	FTVY	162
	orf29.pep	SYVVG	TSTK	TSTK	SIVP	QAPF	SDRW	125
	orf29ng	SYHIG	TSTK	TSTK	INTV	PQAP	FSRW	222

35 The complete length ORF29ng nucleotide sequence <SEQ ID 169> is predicted to encode a protein having amino acid sequence <SEQ ID 170>:

	1	MNLPIQK	FMM	LFAAA	ISLLQ	IPISH	ANGLD	ARLRD	DMQAK	HYEPG	GKYHL
	51	FGNARG	SVKN	RVCAV	QTFDA	TAVGP	ILPIT	HERTG	FEGVI	GYETH	FSGHG
40	101	HEVHSP	FDNH	DSKST	SDFSG	GVDGG	FTVYQ	LHRTG	SEIHP	EDGYD	GPQGG
	151	GYPPPG	GARD	IYSYH	IKGTS	TKTKI	NTVPQ	APFSD	RWLKE	NAGAAS	GFLS
	201	RADEAG	KLIW	ENDPD	KNWRA	NRMDI	RGIV	QGA	VNPFL	TG	FQGLV
	251	DSAVSP	VTYA	AARKT	LQGIH	NLGNL	SPEAQ	LAAAT	ALQDS	AFAVK	DSINS
	301	ARQWAD	AHPN	ITATA	QALAV	AXAAT	TVWVG	GKKV	ELNPAK	WDWV	KNTGYK
	351	KPAARH	MQTV	DGEMAG	GKNKP	LESKN	TVTTN	NFFEN	TGYTE	KVLRQ	ASNGD
45	401	YHGFQ	SVDA	FSENG	TVIQI	VGGDN	IVRHK	LYIPG	SYKKG	DGNFE	YIREA
	451	DGKINH	RFLV	PNQQL	PEK*						

In a second experiment, the following DNA sequence <SEQ ID 171> was identified:

	1	atgAAT	TTCG	CTATT	CAAAA	ATTCAT	GATG	ctgtt	ggcAg	cggca	atatc
	51	gatgct	GCat	ATCCCC	CATTA	GTCAT	GCGAA	CGGTT	TGGAT	GCCCG	TTCG
50	101	GCGAT	GATAT	GCAGG	CAAAA	CACTAC	GAAAC	CGGGT	GGCAA	ATACCA	CTG
	151	TTTGG	TAAATG	CTCGC	GCGCAG	TGTTA	AAAAAT	CGGGT	TTCG	CCGTCC	AAAC
	201	ATTTG	ATGCA	ACTGC	GCGTCG	GCCCC	CATACT	GCCTA	TATACA	CACGA	ACGGA
	251	CAGGA	TTTGA	AGGTG	TTATC	GGCTA	TGAAA	CCCAT	TTTTTC	AGGAC	ACGGA
	301	CACGA	AAGTAC	ACAGT	CCGTT	CGATA	ATCAT	GATTCA	AAAAA	GCACT	TCTGA
55	351	TTTCAG	CGGC	GGCGT	AGACG	GCGGT	TTTAC	CGTTA	ACCAA	CTTCAT	CGGA
	401	CAGGG	TCCGA	AATAC	ATCCC	GCAGC	CGGAT	ATGAC	CGGCC	TCAAG	GCGGC
	451	GGTTAT	CCGG	AACCA	CAAGG	GGCA	AGGGAT	ATATA	CAGCT	ACCAT	ATCAA
	501	AGAACT	TCA	ACCAAA	ACAA	AGATA	AAACAC	TGTTCC	GCAA	GCCCC	TTTTT
	551	CAGAC	CGCTG	GCTAA	AAGAA	AATGC	CGGTG	CCGCT	TCCGG	TTTCT	CAGC
60	601	CGTG	CGGATG	AAGCA	GGAAA	ACTGA	TATGG	GAAA	ACGACC	CCGATA	AAAAA

5	651	TTGGCGGGCT	AACCGTATGG	ATGATATTCG	CGGCATCGTC	CAAGGTGCGG
	701	TTAATCCTTT	TTTAAAGGGT	TTTCAAGGGG	TAGGGATTGG	GGCAATTACA
	751	GACAGTGC GG	TAAAGCCCGT	CACAGATACA	GCCGCTCAGC	AGACTCTACA
	801	AGGTATTAAC	GATTTAGGAA	ATTTAAGTCC	GGAGAGACAA	CTTGCCGCGG
	851	CGAGCCTATT	ACAGGACAGT	GCCTTTGCGG	TAAAAGACGG	CATCAATTCC
10	901	GCCAGACAAT	GGGCTGATGC	CCATCCGAAT	ATAACAGCAA	CAGCCCAAA
	951	TGCCCTTGCC	TAGCAGAGG	CCGCAGGTAC	GTTTGGCGC	GGTAAAAAAG
	1001	TAGAACTTAA	CCCGACCAA	TGGGATTGGG	TTAAAAATAC	CGGCTATAAA
	1051	AAACCTGCTG	CCCGCATAT	GCAGACTGTA	GATGGGGAGA	TGGCAGGGGG
	1101	GAATAGACCG	CCTAAATCTA	TAACGTGCGA	AGGAAAAAGT	AGCTGCGCAA
15	1151	CCTATCCTAA	GTTGGTTAAT	CAGCTAAATG	AGCAAACTT	AAATAACATT
	1201	GCGGCTCAAG	ATCCAAGATT	GAGCTAGCT	ATTCATGAGG	GTAAAAAAAA
	1251	TTTTCCAATA	GGAACTGCAA	CTTATGAAGA	GGCAGATAGA	CTAGGTAAAA
	1301	TTTGGGTTGG	TGAGGGTGCA	AGCAAACTA	GTGGAGGCGG	ATGGTTAAGT
	1351	AGAGATTGGCA	CTCGACAATA	TCGGCCACCA	ACAGAAAAAA	AATCACAATT
	1401	TGCAACTACA	GGTATTCAAG	CAAAATTTTA	AACTTATACT	ATTGATTCAA
	1451	ATGAAAAAAG	AAAAATAAAT	AAAAATGGAC	ATTTAAATAT	TAGGTAA

This encodes a protein having amino acid sequence <SEQ ID 172; ORF29ng-1>:

20	1	MNLPIQKFM	LLAAISMLH	IPISHANGLD	ARLRDDMQAK	HYEPGGKYHL
	51	FGNARGSVKN	RVCAVQTFDA	TAVGPILPIT	HERTGFEGVI	GYETHFSGHG
	101	HEVHSPFDNH	DSKSTDSFSG	GVDGGFTVYQ	LHRTGSEIHP	ADGYDGPQGG
	151	GYPEPQGARD	IYSYHIKGT	TKTKINTVPQ	APFSDRWLKE	NAGAASGFLS
	201	RADEAGKLIW	ENDPDKNWRA	NRMDDIRGIV	QGVNPFITG	FQGVGIGAIT
25	251	DSAVSPVTD	AAQQTLOGIN	DLGNLSPEAQ	LAAASLLQDS	AFAVKDGINS
	301	ARQWADAHPN	ITATAQTALA	VAEAAGTVWR	GKKVELNPTK	WDWVKNTGYK
	351	KPAARMHMTV	DGEMAGGNRP	PKSITSEGKA	NAATYPKLVN	QLNEQNLNNI
	401	AAQDPRLSLA	IHEGKKNFPI	GTATYEEADR	LKGIWVGEBA	RQTSGGGWSL
	451	RDGTRQYRPP	TEKKSQFATT	GIQANFETYT	IDSNEKRNKI	KNGHLNLR*

ORF29ng-1 and ORF29-1 show 86.0% identity in 401 aa overlap:

30		10	20	30	40	50	60
	orf29ng-1.pep	MNLPIQKFMMLAAAI	SMLHIPISHANGLDARLRDDMQAKHYEPGGKYHLFGNARGSVKN				
	orf29-1	MNLPIQKFMMLFAAAISLLQIPISHANGLDARLRDDMQAKHYEPGGKYHLFGNARGSVKK					
35		10	20	30	40	50	60
	orf29ng-1.pep	70	80	90	100	110	120
	orf29-1	RVCAVQTFDATAVGPILPITHERTGFEGVIGYETHFSGHGHEVHSPFDNHDSKSTSDFSG					
40		70	80	90	100	110	120
	orf29ng-1.pep	130	140	150	160	170	180
	orf29-1	GVDGGFTVYQLHRTGSEIH	PDAGYDGPQGGGYPEPQGARDIYSYHIKGTSTTKTINTVPQ				
45		130	140	150	160	170	180
	orf29ng-1.pep	190	200	210	220	230	240
	orf29-1	APFSDRWLKENAGAA	SGFLSRADEAGKLIWENDPDKNWRANRMDDIRGIVQGA				
50		190	200	210	220	230	240
	orf29ng-1.pep	250	260	270	280	290	300
	orf29-1	FQGVGIGAITDS	AVSPVTDTA	AQQT	LQGIN	DGNSPEAQLAAASLLQDS	SAFAVKDGINS
55		250	260	270	280	290	300
	orf29ng-1.pep	310	320	330	340	350	360
	orf29-1	ARQWADAH	PNITATAQTAL	AVAEAA	GTVWRGKKV	ELNPTKWDWVKNTGYKKPAARHMQTV	
60		310	320	330	340	350	360
	orf29ng-1.pep	370	380	390	400	410	420
	orf29-1	AKQWADAH	PNITATAQTAL	SAEAA	GTVWRGKKV	ELNPTKWDWVKNTGYKKPAARHMQTL	
65		370	380	390	400	410	420

```

5
orf29ng-1.pep  DGEMAGGNRPKSI-TSEGKANAATYPKLVNQLNEQNLNNIAAQDPRLSLAIHEGKKNFP
                ||||| |||: || :: :|: :: : ::::
orf29-1         DGEMAGGNKPIKSLPNSAAEKRKQNFEEKFNSNWSSASFDSVHKTLTPNAPGILSPDKVKT
                370          380          390          400          410          420

orf29ng-1.pep  420          430          440          450          460          470          479
                IGTATYEEADRLGKIWVGEGARQTSGGGWLSRDGTRQYRPPTEKKSQFATTGIQANFETY

orf29-1         RYTSLDGKITI IKDNENNYFRIHDNSRKQYLD SNGNAVK TGNLQ GKQAKDY LQQQTHIRN
                430          440          450          460          470          480
10

```

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### 15 Example 21

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 173>:

1 ATGAAAAAAC AAATCACC GC AGCCGTAATG ATGCTGTCTA TGATTGCCCC  
51 CGCAATGGCA AACGCGTTGG ACAATCAGGC ATTTGAAGAC CAAATGTTCC  
101 ACACGCGGGC AGATGCACCG ATGCAG...

20 This corresponds to the amino acid sequence <SEQ ID 174; ORF30>:

1 MKKOITAAVM MLSMIAPAMA NGLDNQAFED QMFHTRADAP MQ..

Further work revealed the complete nucleotide sequence <SEQ ID 175>:

25	1	ATGAAAAAAC	AAATCACCGC	AGCCGTAATG	ATGCTGTCTA	TGATTGCCCC
	51	CGCAATGGCA	AACGGCTTGG	ACAATCAGGC	ATTTGAAGAC	CAAGTGTTCC
	101	ACACGCGGGC	AGATGCACCG	ATGCAGTTGG	CGGAGCTTTC	TCAAAAGGAG
	151	ATGAAGGAGA	CAGAGGGGGC	GTTTCTTCCA	TGGCTATCTC	TGGGTGGTGC
	201	TGCCATTGGT	ATGTGGACAC	AGCATGGTTT	TAGTTATGCA	ACGACAGGCA
30	251	GACCAGCTTC	TGTTAGAGAT	GTTGCTATTG	CTGGCCGATT	AGGCGCAATT
	301	CCTGGTGGTG	TAGGCCCCGC	AGGAAAGGTT	GTTTCCGTTG	CTAAATATGG
	351	ACGTGAGATT	AAAACTGGCA	ATAATATGCG	GATAGCCCTT	TTCGGTAATA
	401	GAACAGGTCA	T CCTATTGGA	AAATTTCCCC	ATTATCATCA	TCGAGTTACG
	451	GATAATACGG	CGAAGACTTT	GCCTGGACAG	GGAATTGGTC	GTCATCGCCC
	501	TTGGGAATCA	AAATCTACGG	ACAGATCATG	GAAAAACCGC	TTCCTAA

This corresponds to the amino acid sequence <SEQ ID 176; ORF30-1>:

35 1 MKKQITAAVM MLSMIAPAMA NGLDNQAFED QVFHTRADAP MQIAELSQKE  
51 MKETEGAFLP LAILGGAIG MWTQHGFSYA TTGRPASVRD VAIAGGLGAI  
101 PGGVGAAGKV VSFAYGREI KIGNNMRIAP FGNRTGHPIG KFPHYHRRVT  
151 DNTGKTLPGO GIGRHRPWES KSTDRSWKNR F\*

Computer analysis of this amino acid sequence gave the following results:

40 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF30 shows 97.6% identity over a 42aa overlap with an ORF (ORF30a) from strain A of *N. meningitidis*:

```

45      orf30.pep      MKKQITAAVMMLSMIAPAMANGLDNQAFEDQMFHTRADAPMQ
                        10      20      30      40
                        |||||
      orf30a          MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKXTXGAFLP
                        10      20      30      40      50      60
      orf30a          LXILGGAAIGMWTQHGFSYATTGRPASVRDVAIAGGLGAIPGXVGAAGKVVSFAKYGREI

```

70 80 90 100 110 120

The complete length ORF30a nucleotide sequence <SEQ ID 177> is:

```

1  ATGAAAAAAC AAATCACCGC AGCCGTAATG ATGCTGTCTA TGATTGCCCC
5  51  CGCAATGGCA AACGGCTTGG ACAATCAGGC ATTTGAAGAC CAAGTGTTC
101 ACACGCGGGC AGATGCACCG ATGCAGTTGG CGGAGCTTTC TCAAAAGGAG
151 ATGAAGGANA CAGNGGGGGC GTTCTTCCA TTGNTATCT TGGGTGGTGC
201 TGCCATTGGT ATGTGGACAC AGCATGGTTT TAGTTATGCA ACGACAGGCA
251 GACCAGCTTC TGTAGAGAT GTTGCTATTG CTGGCGGATT AGGCGCAATT
301 CCTGGTGNTG TAGGCGCCGC AGGAAAGGTT GTTCCCTTG CTAAATATGG
10 351 ACGTGAGATT AAAATCGGCA ATAATATGCG GATAGCCCTT TCGGTAATA
401 GAACAGGTCA TCCTATTGGN AAATTTCCCC ATTATCATCG TCGAGTTACG
451 GATAATACGG GCAAGACTTT GCCTGGACAG GGAATTGGTC GTCATCGCCC
501 TTGGGAATCA AAATCTACGG ACAGATCATG GAAAACCGC TTCTAA

```

This encodes a protein having amino acid sequence <SEQ ID 178>:

```

15 1  MKKQITAAVM MLSMIAPAMA NGLDNQAFED QVFHTRADAP MQLAELSQKE
51 51  MKXTXGAFLP LXILGGAATG MWTQHGFSA TTGRPASVRD VAIAGGLGAI
101 101 PGXVGAAGKV VSFAYGREI KIGNNMRIAP FGNRTGHPIG KFPYHRRVT
151 151 DNTGKTLPGQ GIGRHPWES KSTDRSWKNR F*

```

ORF30a and ORF30-1 show 97.8% identity in 181 aa overlap:

```

20 orf30a.pep      MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKXTXGAFLP 60
    orf30-1        MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKETEGAFLP 60

25 orf30a.pep      LXILGGAATGMWTQHGFSAATTGRPASVRDVAIAGGLGAIPGXVGAAGKVVSFAKYGREI 120
    orf30-1        LAILGGAATGMWTQHGFSAATTGRPASVRDVAIAGGLGAIPGGVGAAGKVVSFAKYGREI 120

30 orf30a.pep      KIGNNMRIAPFGNRTGHPIGKFPYHRRVTDNTGKTLPGQGIGRHPWESKSTDRSWKNR 180
    orf30-1        KIGNNMRIAPFGNRTGHPIGKFPYHRRVTDNTGKTLPGQGIGRHPWESKSTDRSWKNR 180

35 orf30a.pep      FX
    orf30-1        FX

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF30 shows 97.6% identity over a 42aa overlap with a predicted ORF (ORF30.ng) from *N. gonorrhoeae*:

```

40 orf30.pep      MKKQITAAVMMLSMIAPAMANGLDNQAFEDQMFHTRADAPMQ 42
    orf30ng       MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKETEGAFLP 60

```

The complete length ORF30ng nucleotide sequence <SEQ ID 179> is

```

45 1  ATGAAAAAAC AAATCACCGC AGCCGTAATG ATGCTGTCTA TGATCGCCCC
51 51  CGCAATGGCA AACGGATTGG ACAATCAGGC ATTTGAAGAC CAAGTGTTC
101 ACACGCGGGC AGATGCGCCG ATGCAGTTGG CGGAGCTTTC TCAGAAGGAG
151 ATGAAGGAGA CTGAAGGGGC TTTCTTCCA TTGGCTATCT TGGGTGGTGC
201 TGCCATTGGT ATGTGGACAC AGCATGGTTT TAGTTATGCA ACGACAGGCA
251 GACCAGCTTC TGTAGAGAT GTTGCTGGCG GATTAGGCGC AATTCCTGGT
301 GATGTAGGTG CTGCAGGAAA GGTGTTTCC TTTGCTAAAT ATGGACGTGA
50 351 GATTAAAATC GGCAATAATA TGCGGATAGC CCCTTTCGGT AATAGAACAG
401 GTCATCCTAT TGGAAAATTT CCCCATTATC ATCGTCGAGT TACGGATAAT
451 ACGGGCAAGA CTTTGCTTGG ACAGGGAATT GGTGCTCATC GCCCTTGGGA
501 ATCAAAATCT ACGGACAGAT CATGGAAAAA CCGCTTCTAA

```

This encodes a protein having amino acid sequence <SEQ ID 180>:

```

55 1  MKKQITAAVM MLSMIAPAMA NGLDNQAFED QVFHTRADAP MQLAELSQKE

```



51 MKETEGAFLP LAILGGAAG MWTQHGFSA TTGRPASVRD VAGGLGAIPG  
 101 DVGAGKVVS FAKYGREIKI GNNMRIAPFG NRTGHPGKF PHYHRRVTDN  
 151 TGKTLPGQGI GRHRPWESKS TDRSWKNRF\*

ORF30ng and ORF30-1 show 98.3% identity in 181 aa overlap:

5		10	20	30	40	50	60
	orf30ng.pep	MKKQITA	AVMMLSM	IAPAMAN	GLDNQAF	EDQVFH	TRADAPM
	orf30-1	MKKQITA	AVMMLSM	IAPAMAN	GLDNQAF	EDQVFH	TRADAPM
10		70	80	90	100	110	
	orf30ng.pep	LAILGGA	AIGMWTQ	HGFSYAT	TGRPASV	RDVAGGL	GAIPGDV
	orf30-1	LAILGGA	AIGMWTQ	HGFSYAT	TGRPASV	RDVAGGL	GAIPGDV
15		120	130	140	150	160	170
	orf30ng.pep	KIGNNM	RIAPFGN	RTPHPIG	KFPHYHR	RTDNTG	KTLPGQG
	orf30-1	KIGNNM	RIAPFGN	RTPHPIG	KFPHYHR	RTDNTG	KTLPGQG
20		130	140	150	160	170	180
	orf30ng.pep	IGRHRP	WESKST	DRSWKN	RFX		
	orf30-1	IGRHRP	WESKST	DRSWKN	RFX		

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

## Example 22

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 181>:

1 ATGAATAAAA CTCTCTATCG TGTAATTTTC AACCGCAAAC GTGGGGCTGT  
 51 GrTAGCCGTT GCTGAACTA CCAAGCGCGA AGGTAAAAGC TGTGCCGATA  
 101 GTGATTTCAGG CAGCGCTCAT GTGAAATCTG TTCCTTTTGG TACTACTCAT  
 151 GCACCTGTTT GTg.CGTtAc AAATATCTTT TCTTTTCTT TATTGGGCTT  
 201 TTCTTTATGT TTGGCTGTAG GtacGGyCAA TATTGCTTTT GCTGATGGCA  
 251 TT..

This corresponds to the amino acid sequence <SEQ ID 182; ORF31>:

1 MNKTLYRVIF NRKRGAVXAV AETTKREGKS CADSDSGSAH VKSVFPGTTH  
 51 APVCXVTNIF SFSLLGFSLC LAVGTXNIAF ADGI..

Further work revealed a further partial nucleotide sequence <SEQ ID 183>:

1 ATGAATAAAA CTCTCTATCG TGTAATTTTC AACCGCAAAC GTGGGGCTGT  
 51 GGTAGCCGTT GCTGAACTA CCAAGCGCGA AGGTAAAAGC TGTGCCGATA  
 101 GTGATTTCAGG CAGCGCTCAT GTGAAATCTG TTCCTTTTGG TACTACTCAT  
 151 GCACCTGTTT GTCGTTCAAA TATCTTTTCT TTTTCTTTAT TGGGCTTTTC  
 201 TTTATGTTTG GCTGTAGGTA CGGCCAATAT TGCTTTTGCT GATGGCATT..

This corresponds to the amino acid sequence <SEQ ID 184; ORF31-1>:

1 MNKTLYRVIF NRKRGAVVAV AETTKREGKS CADSDSGSAH VKSVFPGTTH  
 51 APVCRSNIFS FSLGFSLCL AVGTANIAFA DGI..

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.gonorrhoeae*

ORF31 shows 76.2% identity over a 84aa overlap with a predicted ORF (ORF31.ng) from *N.*

*gonorrhoeae*:

```

5      orf31.pep      MNKTLRVIFNKRKGA VXAETTKREGKSCADSDSGSAHVKSVPFGTTHAPVCXVTNIF      60
      orf31ng        MNKTLRVIFNKRKGA VXAETTKREGKSCADSDSGSVYVKSVSFIPTH-----SKAF      54

      orf31.pep      SFSLLGFSLCLAVGTXNIAFADGI      84
10     orf31ng        CFSALGFSLCLALGTVNIAFADGIITDKAAPKTQQATILQTGNGIPQVNIQTPTSAGVSV      114

```

The complete length ORF31ng nucleotide sequence <SEQ ID 185> is:

```

1      ATGAACAAAA  CCCTCTATCG  TGTGATTTTC  AACCGCAAAC  GCGGTGCTGT
51     GGTAGCTGTT  GCCGAAACCA  CCAAGCGCGA  AGGTAAAAGC  TGTGCCGATA
101    GTGGTTCGGG   CAGCGTTTAT  GTGAAATCCG  TTTCTTTTAT  TCCTACTCAT
151    TCCAAAGCCT  TTTGTTTTTC  TGCATTAGGC  TTTTCTTTAT  GTTTGGCTTT
201    GGGTACGGTC  AATATTGCTT  TTGCTGACGG  CATTATTACT  GATAAAGCTG
251    CTCCTAAAAC  CCAACAAGCC  ACGATTCTGC  AAACAGGTaa  cGGCATACCG
301    CAAGTCAATA  TTCAAACCCC  TACTTCGGCA  GGGGTTTCTG  TTAATCAATA
351    TGCCCAAGTT  GATGTGGGTA  ATCGCGGGGC  GATTTTAAAC  AACAGTCGCA
401    GCAACACCCA  AACACAGCTA  GCGGTTGGA  TTCAAGGCAA  TCCTTGCTTG
451    ACAAGGGGCG  AAGCACGTGT  GGTGTAAAC  CAAATCAACA  GCAGCCATCC
501    TTCACAACCT  AATGGCTATA  TTGAAGTGG  TGGACGACGT  GCAGAAGTCG
551    TTATTGCCAA  TCCGGCAGGG  ATTGCAGTCA  ATGGTGGTGG  TTTTATCAAT
601    GCTTCCCGTG  CCACTTTGAC  GACAGGCCAA  CCGCAATATC  AAGCAGGAGA
25     651    CTTTAGCGGC  TTAAAGATAA  GGCAAGGCAA  TGCTGTAATC  GCCGGACACG
701    GTTTGGATGC  CCGTGATACC  GATTTACAC  GTATTCTGT  ATGCCAACAA
751    AATCACCTTG  ATCAGTACGG  CCGAACAGC  AGGCATTCGT  AA

```

This encodes a protein having amino acid sequence <SEQ ID 186>:

```

1      MNKTLRVIF  NRKRGAVVAV  AETTKREGKS  CADSGSGSVY  VKSVSFIPTH
51     SKAFCFSALG  FSLCLALGTV  NIAFADGIIT  DKAAPKTQQA  TILQTGNGIP
101    QVNIQTPTSA  GVSVNQYQF  DVGNRGAILN  NSRSNTQTQL  GGWIQGNPWL
151    TRGEARVVVN  QINSSHPSQL  NGYIEVGRR  AEVVIANPAG  IAVNGGGFIN
201    ASRATLTGQ  PQYQAGDFSG  FKIRQGNNA  AGHGLDARDT  DFTRILVCQQ
251    NHLDQYGRS  RHS*

```

35 This gonococcal protein shares 50% identity over a 149aa overlap with the pore-forming hemolysins-like HecA protein from *Erwinia chrysanthemi* (accession number L39897):

```

40     orf31ng      96  GNGIPQVNIQTPTSAGVSVNQAQFDVGNRGAILNNSRSN-TQTQLGGWIQGNPWLTRGE 154
      HecA         45  GNGVPVVNIATPDASGLSHNRYHDFNVDNRGLILNNGTARLTSPQLGGLIQNNPNLNGRA 104

      Orf31ng      155  ARVVVNQINSSHPSQLNGYIEVGRRRAEVVIANPAGIAVNGGGFINASRATLTGQPQYQ 214
      HecA         105  AAAILNEVVSPNRSRLAGYLEVAGQAANVVVANPYGITCSGCGFLNTPRLTLTTGTPTQFD 164

45     Orf31ng      215  -AGDFSGFKIRQGNNAVIAGHGLDARDTDF 242
      HecA         165  AAGGLSGLDVRGGDILIDGAGLDASRSDY 193

```

Furthermore, ORF31ng and ORF31-1 show 79.5% identity in 83 aa overlap:

```

50     orf31-1.pep  MNKTLRVIFNKRKGA VXAETTKREGKSCADSDSGSAHVKSVPFGTTHAPVCRSNI FS
      orf31ng       MNKTLRVIFNKRKGA VXAETTKREGKSCADSDSGSVYVKSVSFIPTH-----SKAFC
      10          20          30          40          50          60

55     orf31-1.pep  FSLLGFSLCLAVGTANIAFADGI

```

orf31ng                    || |||||:|||||  
 FSALGFSLLCLALGTVNIAFADGIITDKAAPKTQQATILQTGNGIPQVNIQTPTSAGVSVN  
                               60              70              80              90              100             110

On this basis, including the homology with hemolysins, and also with adhesins, it is predicted that  
 5 the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens  
 for vaccines or diagnostics, or for raising antibodies.

**Example 23**

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 187>:

10                    1 ATGAATACTC CTCCTTTTGT CTGTTGGATT TTTTGCAAGG TCATCGACAA  
                       51 TTTCCGGCGAC ATCGGCGTTT CGTGGCGGCT CGCCCGTGTT TTGCACCGCG  
                      101 AACTCGGTTG GCAGGTGCAT TTGTGGACGG ACGATGTGTC CGCCTTGCGT  
                      151 GCGCTTTGCC CTGATTTGCC CGATGTTCCC TCGGTTTCATC AGGATATTCA  
                      201 TGTCGCGACT TGGCATTCCG ATGCGGCAGA TATTGATACC GCG..

This corresponds to the amino acid sequence <SEQ ID 188; ORF32>:

15                    1 MNTPPFVCWI FCKVIDNFGD IGVSWRLARV LHRELGWQVH LWTDDVSALR  
                      51 ALCPDLPDVP CVHQDIHVRT WHSDAADIDT A..

Further work revealed the complete nucleotide sequence <SEQ ID 189>:

20                    1 ATGAATACTC CTCCTTTTGT CTGTTGGATT TTTTGCAAGG TCATCGACAA  
                      51 TTTCCGGCGAC ATCGGCGTTT CGTGGCGGCT CGCCCGTGTT TTGCACCGCG  
                      101 AACTCGGTTG GCAGGTGCAT TTGTGGACGG ACGATGTGTC CGCCTTGCGT  
                      151 GCGCTTTGCC CTGATTTGCC CGATGTTCCC TCGGTTTCATC AGGATATTCA  
                      201 TGTCGCGACT TGGCATTCCG ATGCGGCAGA TATTGATACC GCGCCTGTTC  
                      251 CCGATGTGCT CATCGAAACT TTTGCTGCG ACCTGCCCGA AAATGTGCTG  
                      301 CACATTATCC GCCGACACAA GCCGCTTTGG CTGAATTGGG AATATTTGAG  
                      351 CGCGGAGGAA AGCAATGAAA GGCTGCATCT GATGCCTTCG CCGCAGGAGG  
                      401 GTGTTCAAAA ATATTTTGG TTTATGGGTT TCAGCGAAAA AAGCGGCGGG  
                      451 TTGATACGCG AACGTGATTA CTGCGAAGCC GTCCGTTTCG ATACTGAAGC  
                      501 CCTGCGAGAG CGGCTGATGC TGCCCCGAAA AACGCCTCC GAATGGCTGC  
                      551 TTTTCGGCTA TCGGAGCGAT GTTTGGGCAA AGTGGCTGGA AATGTGGCGA  
                      601 CAGGCAGGCA GCCCGATGAC ACTGTTGCTG GCGGGGACGC AAATCATCGA  
                      651 CAGCCTCAAA CAAAGCGGCG TTATTCGCA AGATGCCCTG CAAAACGACG  
                      701 GCGATGTTT TCAGACGGCA TCCGTCCGCC TCGTCAAAAT CCTTTTCGTG  
                      751 CCGCAACAGG ACTTCGACCA ACTGCTGCAC CTTGCCGACT GCGCCGTCAT  
                      801 CCGCGCGGAA GACAGTTTCG TCGCGGCCA GCTTGCGGGC AAACCTTCT  
                      851 TTTGGCACAT CTACCCGCAA GACGAGAATG TCCATCTCGA CAACTCCAC  
                      901 GCCTTTTGGG ATAAGGCACA CGGTTTCTAC ACGCCCGAAA CCGTGTGCGC  
                      951 ACACCGCCGT CTTTCGGACG ACCTCAACGG CGGAGAGGCT TTATCCGCAA  
                      1001 CACAACGCCT CGAATGTTGG CAAACCTGC AACAACATCA AAACGGCTGG  
                      1051 CGGCAAGGCG CGGAGGATTG GAGCCGTTAT CTTTTCGGGC AGCCGTCAGC  
                      1101 TCCTGAAAAA CTCGCTGCCT TTGTTTCAA GCATCAAAA ATACGCTAG

This corresponds to the amino acid sequence <SEQ ID 190; ORF32-1>:

45                    1 MNTPPFVCWI FCKVIDNFGD IGVSWRLARV LHRELGWQVH LWTDDVSALR  
                      51 ALCPDLPDVP CVHQDIHVRT WHSDAADIDT APVPDVVIET FACDLPENVL  
                      101 HIIRRHKPLW LNWEYLSAEE SNERLHLMPS PQEGVQKYFW FMGFSEKSGG  
                      151 LIRERDYCEA VRFDEALRE RLMLPEKNAS EWLLFGYRSD VWAKWLEMR  
                      201 QAGSPMTLLL AGTQIIDLK QSGVIPQDAL QNDGDVFQTA SVRLVKIPFV  
                      251 PQQDFDQLLH LADCAVIRGE DSFVRAQLAG KPFFWHIYPQ DENVHLDKLH  
                      301 AFWDKAHGFY TPETVSAHRR LSDDLNGGEA LSATQRLCW QTLQQHQNGW  
                      351 RQGAEDWSRY LFGQPSAPEK LAAFVSKHQK IR\*w

50 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)ORF32 shows 93.8% identity over a 81aa overlap with an ORF (ORF32a) from strain A of *N.**meningitidis*:

```

      10      20      30      40      50      60
5    orf32.pep  MNTPPFVCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDVVSALRALCPDLPDVP
      |||||
    orf32a      MNTPPFSAGXFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDVVSALRALCPDLPDVX
      10      20      30      40      50      60

      70      80
10   orf32.pep  CVHQDIHVRTWHSDAADIDTA
      |||||
    orf32a      CVHQDIHVRTWHSDAADIDTAPVXDVIETFACDLPENVLHIIRRHKPLWLXWEYLSAEX
      70      80      90      100     110     120
```

15 The complete length ORF32a nucleotide sequence &lt;SEQ ID 191&gt; is:

```

1  ATGAATACTC CTCCTTTTTC TGCTGGANTT TTTTGCAAGG TCATCGACAA
51 TTTCGGCGAC ATCGGCGTTT CGTGGCGGCT TGCCCGTGTT TTGCACCGCG
101 AACTCGGTTG GCAGGTGCAT TTGTGGACGG ACATGTGTGC CGCCTTGCGT
151 GCGCTTTGCC CTGATTTGCC CGATGTTTNC TGCGTTCATC AGGATATTCA
201 TGTCGCACT TGGCATTCCG ATGCGGCAGA TATTGATACC GCGCCTGTTC
251 NCGATGTCGT CATCGAAACT TTTGCCTGCG ACCTGCCCGA AAATGTGCTG
301 CACATCATCC GCCGACACAA GCCGCTTTGG CTGAANTGGG AATATTTGAG
351 CGCGGAGGAN AGCAATGAAA GGCTGCACNT GATGCCTTCG CCGCAGGAGA
401 GTGTTTCAAA ATANTTTTGG TTATGGGTT TCAGCGAANN NAGCGCGGGA
25 451 CTGATACGCG AACGCGATTA CTGCGAAGCC GTCCGTTTCG ATAGCGGAGC
501 CTTCGCGCAAG AGGCTGATGC TTCCCGAAAA AAACGNCCCC GAATGGCTGC
551 TTTTCGGCTA TCGGAGCGAT GTTGGGCAA AGTGGCTGGA AATGTGGCGA
601 CAGGCAGGCA GTCCGTTGAC ACTTTTGCTG GCNNGGGGCGC ANATTATCGA
651 CAGCCTCAAA CAAAACGGCG TTATTCGCGA AGATGCCCTG CAAAACGACG
30 701 GCGATGTTTT TCAGACGGCA TCCGTCCGCC TCGTCAAAAT CCCTTTCGTG
751 CCGCAACAGG ACTTCGACAA ACTGCTGCAC CTGCGCGACT GCGCCGTCAT
801 CCGCGGCGAA GACAGTTTCG TGCGCGCCCA GCTTGCGGGC AAACCTTCT
851 TTTGGCACAT CTACCCGCAA GATGAGAAATG TCCATCTCGA CAAACTCCAC
901 GCCTTTTGGG ATAAGGCACA CGGTTTCTAC ACGCCCGAAA CCGCATCGGC
35 951 ACACCGCCGC CTTTCAGACG ACCTCAACGG CGGAGAGGCT TTATCCGCAA
1001 CACAACGCCT CGAATGTTGG CAAATCCTGC AACAACATCA AAACGGCTGG
1051 CGGCAAGGCG CGGAGGATTG GAGCCGTTAT CTTTTTGGGC AGCCTTCCGC
1101 ATCCGAAAAA CTGCGCGCCT TTGTTTCAAA GCATCAAAAA ATACGCTAG
```

This encodes a protein having amino acid sequence &lt;SEQ ID 192&gt;:

```

40 1  MNTPPFSAGX FCKVIDNFGD IGVSWRLARV LHRELGWQVH LWTDDVSALR
51 51  ALCPDLPDVX CVHQDIHVRT WHSDAADIDT APVXDVIET FACDLPENVL
101 101 HIIRRHKPLW LXWEYLSAEX SNERLHXMP S QESVXXKFEW FMGFSEKSGG
151 151 LIRERDYCEA VRFDSGALRK RLMLPEKNXP EWLLFGYRSD VWAKWLEMWR
201 201 QAGSPLTLLL AGAXIIDS LK QNGVIPQDAL QNDGDV FQTA SVRLVKIPFV
45 251 PQQDFDKLLH LADCAVIRGE DSFVRAQLAG KPFFWHIYPQ DENVHLDKLH
301 301 AFWDKAHGFY TPETASAHRR LSDDLNGGEA LSATQRLECW QILQQHQNGW
351 351 RQGAEDWSRY LFGQPSASEK LAAFVSKHQK IR*
```

ORF32a and ORF32-1 show 93.2% identity in 382 aa overlap:

```

      10      20      30      40      50      60
50   orf32-1.pep MNTPPFVCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDVVSALRALCPDLPDVP
      |||||
    orf32a      MNTPPFSAGXFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDVVSALRALCPDLPDVX
      10      20      30      40      50      60

      70      80      90      100     110     120
55   orf32-1.pep CVHQDIHVRTWHSDAADIDTAPVDPVVIETFACDLPENVLHIIRRHKPLWLXWEYLSAEE
      |||||
    orf32a      CVHQDIHVRTWHSDAADIDTAPVXDVIETFACDLPENVLHIIRRHKPLWLXWEYLSAEX
      70      80      90      100     110     120
60
```

		130	140	150	160	170	180
	orf32-1.pep	SNERLHLMPS	QEGVQKYFW	FMGFSEKSG	LIRERDYCE	AVRFDTEAL	RERLMLPEKNAS
	orf32a	SNERLHXMPS	QESVXKXFW	FMGFSEXSG	LIRERDYCE	AVRFDGALR	KRLMLPEKNXP
5		130	140	150	160	170	180
	orf32-1.pep	EWLLFGYRS	DVWAKWLEM	WRQAGSPMT	LLLAGTQI	IDSLKQSGV	IPQDALQNDG
10	orf32a	EWLLFGYRS	DVWAKWLEM	WRQAGSPMT	LLLAGAXI	IDSLKQNGV	IPQDALQNDG
		190	200	210	220	230	240
	orf32-1.pep	SVRLVKIPF	VQQDFDQL	LHLADCAV	IRGEDSFV	RAQLAGKPF	FWHIYPQDEN
15	orf32a	SVRLVKIPF	VQQDFDKL	LHLADCAV	IRGEDSFV	RAQLAGKPF	FWHIYPQDEN
		250	260	270	280	290	300
	orf32-1.pep	AFWDKAHGF	YTPETVSA	HRLSDDL	NGGEALS	SATQRL	ECWQTLQQH
20	orf32a	AFWDKAHGF	YTPETASA	HRLSDDL	NGGEALS	SATQRL	ECWQTLQQH
		310	320	330	340	350	360
	orf32-1.pep	LFGQPSA	PEKLA	AFVSKH	QKIRX		
25	orf32a	LFGQPSA	PEKLA	AFVSKH	QKIRX		
		370	380				

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF32 shows 95.1% identity over a 82aa overlap with a predicted ORF (ORF32.ng) from *N. gonorrhoeae*:

35	orf32.pep	MNTPPF-V	WCWIFCK	VIDNFGD	IGVSWRL	ARVLHREL	GWQVHLWT	DDVSALRAL	CPDLP	57
	orf32ng	MVMNTYA	FPVCWIF	CKVIDN	FGDIGV	SWRLARV	LHRELGW	QVHLWTDD	VSAALRAL	CPDLP
	orf32.pep	DVPCVHQ	DIHVRT	WHS	DAADIT	A				81
40	orf32ng	DVPFVHQ	DIHVRT	WHS	DAADIT	APVPDA	VIETFAC	DLPENVL	NIIRRHK	PLWLNWE

An ORF32ng nucleotide sequence <SEQ ID 193> was predicted to encode a protein having amino acid sequence <SEQ ID 194>:

	1	MVMNTYA	FPV	CWIFCK	VIDN	FGDIGV	SWRL	ARVLHREL	GW	QVHLWT	DDVS
45	51	ALRALCP	DLP	DVPFVHQ	DIH	VRTWHS	DAAD	IDTAPV	PD	AVIETFAC	DLPE
	101	NVLNIIR	RHK	PLWLNWE	YLS	AESNER	LHL	MPSPQEG	VQK	YFWFMGF	SEK
	151	SGGLIR	ERDY	REAVRFD	TEA	LRRRLV	LPEK	NAPEWLL	FGY	RGDVWAK	WLD
	201	MWQQAG	SLMT	LLLAGA	QIID	SLKQSG	VIPQ	NALQNE	GGVF	QTASVRL	VKI
	251	PFVPQQ	DFDK	LLHLAD	CAVI	RGEDSF	VRTQ	LAGKPF	FWHI	YPQDEN	VHLD
	301	KLHAFW	DKAY	GFYTPET	ASV	HRLSDDL	NG	GEALSAT	QRL	ECGVL*	
50	1	ATGAAT	ACAT	ACGCTT	TTTC	TGTCTG	TGG	ATTTT	TGCA	AGGTCAT	CGA
	51	CAATT	TCGGC	GACATC	GGCG	TTTCGT	GGCG	GCTCGC	CCCGT	GTTTGC	ACC
	101	GCGA	ACTCG	G	TTGGC	AGGTG	CATTGT	GGA	CGGACG	ACGT	GTC
	151	CGCG	GCTTT	GTCCCG	ATT	GCCCGA	TGTT	CCCTTC	GTT	ATCAGG	ATAT
55	201	TCATG	TCCGC	ACTTGG	CATT	CCGATG	CGGC	AGACAT	TGAT	ACCGCG	CCCG
	251	TTCCG	GATGC	CGTTAT	CGAA	ACTTTT	GCCT	GCGAC	CTGCC	CGAAA	ATGTG
	301	CTGA	ACATCA	TCCGCC	GACA	CAAACG	CCTT	TGGCTG	GAAT	GGGAAT	ATTT
	351	GAGCG	CGGAG	GAAAGCA	ATG	AAAGG	CTGCA	CCTGAT	GCCT	TCGCCG	CAGG
	401	AGGCG	GTTCA	AAAATAT	TTT	TGGTTT	TATG	GTTTC	AGCGA	AAAAG	CGGC
60	451	GGGTT	GATAC	GCGAAC	CGCA	TTACCG	CGAA	GCCGT	CCGTT	TCGATA	ACCGA

5 501 AGCCCTGCGC CGGCGGCTGG TGCTGCCC GA AAAAACGCC CCCGAATGGC  
 551 TGCTTTTCGG CTATCGGGGC GATGTTTGGG CAAAGTGGCT GGACATGTGG  
 601 CAACAGGCAG GCAGCCTGAT GACCCCTACTG CTGGCGGGGG CGCAAATTAT  
 651 CGACAGCCTC AAACAAGCG GCGTTATTCC GCAAACGCC CTGCAAAAtg  
 701 aaggcgGTGT CTTTCagacG gcatccgTcC gccttGTCAA AAtcCCGTC  
 751 GTGcCGCAAC AGGAcTTCGA CAAATTGCTG CAcctgcCG ACTGCGCCGT  
 801 GATACGCGGC GAAGACAGTT TCGTGCCTAC CCAGCTTGCC GGAAAACCTT  
 851 TTTTTTGGCA CATCTACCCG CAAGACGAGA ATGTCCATCT CGACAAACTC  
 901 CACGCCTTTT GGGATAAGGC ATACGGCTTC TACACGCCCC AAACCGCATC  
 10 951 GGTGCACCGC CTCCTTTTCGG ACGACCTCAA CGGCGGAGAG GCTTTATCCG  
 1001 CAACACAACG CCTCGAATGT TGGCAAACCC TGCAACAACA TCAAACGGC  
 1051 TGGCGGCAAG GCGCGGAGGA TTGGAGCCGT TATCTTTTCG GGCAGCCTTC  
 1101 CGCATCCGAA AAACTCGCCG CCTTTGTTTC AAAGCATCAA AAAATACGCT  
 1151 AG

15 This encodes a protein having amino acid sequence <SEQ ID 196; ORF32ng-1>:

1 MNTYAFPVCW IFCKVIDNFG DIGVSWRLAR VLHRELGWQV HLWTDVVSAL  
 51 RALCPDLPDV PFVHQDIHVR TWSDAADID TAPVPDAVIE TFACDLPENV  
 101 LNIIRRHKPL WLNWEYLSAE ESNERLHLM SPQEGVQKYF WFMGFSEKSG  
 151 GLIRERDYRE AVRFDEALR RRLVLPEKNA PEWLLFGYRG DVWAKWLDMM  
 20 201 QQAGSLMTLL LAGAIIIDSL KQSGVIPQNA LQNEGGVFQT ASVRLVKIPF  
 251 VPQQDFDKLL HLADCAVIRG EDSFVRTQLA GKPFWHIYP QDENVHLDKL  
 301 HAFWDKAYGF YTPETASVHR LLSDDLNGGE ALSATQRLC WQTLQQHQNG  
 351 WRQGAEDWSR YLFGQPSASE KLAAFVSKHQ KIR\*

ORF32ng-1 and ORF32-1 show 93.5% identity in 383 aa overlap:

25 orf32-1.pep 10 20 30 40 50 59  
 MNTPPF-VCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDVVSALRALCPDLPDV  
 orf32ng-1 MNTYAFPVCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDVVSALRALCPDLPDV  
 30 60 70 80 90 100 110 119  
 orf32-1.pep PCVHQDIHVRTWSDAADIDTAPVPDVVIETFACDLPENVLHNIIRRHKPLWLNWEYLSAE  
 orf32ng-1 PFVHQDIHVRTWSDAADIDTAPVPDAVIETFACDLPENVLNIIRRHKPLWLNWEYLSAE  
 35 70 80 90 100 110 120  
 120 130 140 150 160 170 179  
 orf32-1.pep ESNERLHLMSPQEGVQKYFWFMGFSEKSGGLIRERDYCEAVRFDEALRRLMLPEKNA  
 orf32ng-1 ESNERLHLMSPQEGVQKYFWFMGFSEKSGGLIRERDYREAVRFDEALRRLVLPEKNA  
 40 130 140 150 160 170 180  
 180 190 200 210 220 230 239  
 orf32-1.pep SEWLLFGYRSDVWAKWLEMWRQAGSPMTLLLAGTQIIIDSLKQSGVIPQDALQNDGDVFQT  
 orf32ng-1 PEWLLFGYRGDVWAKWLDMMQQAGSLMTLLLAGAQIIDSILKQSGVIPQNALQNEGGVFQT  
 45 190 200 210 220 230 240  
 240 250 260 270 280 290 299  
 orf32-1.pep ASVRLVKIPFVPQQDFDKLLHLADCAVIRGEDSFVRAQLAGKPFWHIYPQDENVHLDKL  
 orf32ng-1 ASVRLVKIPFVPQQDFDKLLHLADCAVIRGEDSFVRTQLAGKPFWHIYPQDENVHLDKL  
 50 250 260 270 280 290 300  
 300 310 320 330 340 350 359  
 orf32-1.pep HAFWDKAHGFTYTPETVSAHRRLLSDDLNGGEALSATQRLCQWTLQQHQNGWRQGAEDWSR  
 orf32ng-1 HAFWDKAYGFTYTPETASVHRRLLSDDLNGGEALSATQRLCQWTLQQHQNGWRQGAEDWSR  
 55 310 320 330 340 350 360  
 360 370 380  
 orf32-1.pep YLFGQPSAPEKLAAFVSKHQKIRX  
 orf32ng-1 YLFGQPSASEKLAAFVSKHQKIRX  
 65 370 380

On this basis, including the RGD sequence in the gonococcal protein, characteristic of adhesins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF32-1 (42kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 7A shows the results of affinity purification of the His-fusion protein, and Figure 7B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for ELISA, giving a positive result. These experiments confirm that ORF32-1 is a surface-exposed protein, and that it is a useful immunogen.

#### Example 24

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 197>:

```

1  .TTGTTCTGCG GTGTNAAAGT GGGGCGTTTT TTCAGCAGTC CGGCGACGTG
51  GTTTCGGGNC AAAGACCCTG TAAATCAGGC GGTGTGCGG CTGTATNCGG
101 ACGAGTGGCG GCA.ACTTCG GTACGTGGA AAATAGNCGC AACGTCGCAC
151 AGCCTGTGGC TCTGCACGCT GCTCGGAATG CTGGTGTGCG TATTGTTGCT
201 GCTTTTGGTG CGGCAATATA CGTTCAACTG GGAAAGCACG CTGTTGAGCA
251 ATGCCGCTTC GGTACGCGCG GTGGAAATGT TGGCATGGCT GCCGTCGAAA
301 CTCGGTTTCC CTGTCCCCGA TCGCGGCTCG GTCATCGAAG GCCGCTGAA
351 CGGCAATATT GCCGATGCGC GGGCTTGGTC GGGGCTGCTG GTCGNCAGTA
401 TCGCCTGCTA NGGCATCCTG CCGCGCCTG..

```

This corresponds to the amino acid sequence <SEQ ID 198; ORF33>:

```

1  .LFLRVKVGRF FSSPATWFRX KDPVNQAVLR LYXDEWRXTS VRWKIXATSH
51  SLWLCTLLGM LVSLLLLLV RQYTFNWEST LLSNAASVRA VEWLAWLPSC
101 LGFPVPDARS VIEGRLNGNI ADARAWGSLV VXSIACXGIL PRL..

```

Further work revealed the complete nucleotide sequence <SEQ ID 199>:

```

1  ATGTTGAATC CATCCCGAAA ACTGGTTGAG CTGGTCCGTA TTTTGGACGA
51  AGGCGGTTTT ATTTTCAGCG GCGATCCCGT ACAGGCGACG GAGGCTTTGC
101 GCCGCGTGGG CCGCAGTACG GAGGAAAAAA TCATCCGTCG GCGGAGATG
151 ATTGACAGGA ACCGTATGCT GCGGGAGACG TTGGAACGTG TGCCTGCGGG
201 GTCGTCTGCG TTGTGGGTGG TGGCGGCGAC GTTTCGATTT TTTACCGGTT
251 TTTCAGTCAC TTATCTTCTA ATGGACAATC AGGGTCTGAA TTTCTTTTTC
301 GTTTTGGCGG GCGTGTGGG CATGAATACG CTGATGCTGG CAGTATGGTT
351 GGCAATGTTG TTCCTGCGTG TGAAAGTGGG GCGTTTTTTC AGCAGTCCGG
401 CGACGTGGTT TCGGGGCAAA GACCCTGTAA ATCAGGCGGT GTTGCGGCTG
451 TATGCGGACG AGTGGCGGCA ACCTTCGGTA CGTTGGAAAA TAGGCGCAAC
501 GTCGACAGC CTGTGGCTCT GCACGCTGCT CGGAATGCTG GTGTCGGTAT
551 TGTTGCTGCT TTTGGTGGG CAATATACGT TCAACTGGGA AAGCAGCTG
601 TTGAGCAATG CCGCTTCGGT ACGCGCGGTG GAAATGTTGG CATGGCTGCC
651 GTCGAAACTC GGTTCCTCTG TCCCCGATGC GCGGGCGGTC ATCGAAGGCC
701 GTCTGAACGG CAATATTGCC GATGCGCGGG CTGGTGGGG GCTGCTGGTC
751 GGCAGTATCG CCTGCTACGG CATCCTGCGG CGCCTGCTGG CTGGGGTAGT
801 GTGTAAAATC CTTTGA AAA CAAGCGAAAA CGGATTGGAT TTGGAAGAGC
851 CCTATTATCA GCGGCTCATC CGCGCTGGC AGAACAAAAT CACCGATGCG
901 GATACGCGTC GGGAAACCGT GTCCGCGGTT TCACCGAAAA TCATCTTGAA
951 CGATGCGCGG AAATGGGCGG TCATGCTGGA GACCGAGTGG CAGGACGGCG
1001 AATGGTTTGA GGGCAGGCTG GCGCAGGAAT GGCTGGATAA GGGCGTTGCC
1051 ACCAATCGGG AACAGGTTGC CGCGCTGGAG ACAGAGCTGA AGCAGAAACC
1101 GGCGCAACTG CTTATCGGCG TCGCGGCCA AACTGTGCCG GACCGCGGCG
1151 TGTTCGGGCA GATTGTCCGA CTCTCGGAAG CGGCGCAGGG CGGCGCGGTG

```

1201 GTGCAGCTTT TGGCGGAACA GGGGCTTTCA GACGACCTTT CGGAAAAGCT  
 1251 GGAACATTGG CGTAACGCGC TGGCCGAATG CGGCGCGGCG TGGCTTGAGC  
 1301 CTGACAGGGC GGCGCAGGAA GGGCGTTTGA AAGACCAATA A

This corresponds to the amino acid sequence <SEQ ID 200; ORF33-1>:

5 1 MLNPSRKLVE LVRILDEGGF IFSGDPVQAT EALRRVDGST EEKIIRRAEM  
 51 IDRNRMLRET LERVVAGSEW LWVVAATFAF FTGFSVTYLL MDNQGLNFFL  
 101 VLAGVLGMNT LMLAVWLAML FLRVKVGRRF SSPATWFRGK DPNQAVLRL  
 151 YADEWRQPSV RWKIGATSHS LWLCTLLGML VSVLLLLLVLR QYTFNWESTL  
 201 LSNAASVRV EMLAWLPSKL GFPVPDARAV IEGRNLGNIA DARAWSGLLV  
 10 251 GSIACYGILP RLLAWVVCKI LLKTSENGLD LEKPYQAVI RRWQNKITDA  
 301 DTRRETQSAV SPKILNDAP KWAVMLETEW QDGEWFEGRL AQEWLDKGVA  
 351 TNREQVALE TELKQKPAQL LIGVRAQTVP DRGVLRQIVR LSEAAQGGAV  
 401 VQLLAEQGLS DDLSEKLEHW RNALAECGAA WLEPDRAAQE GRLKQD\*

Computer analysis of this amino acid sequence gave the following results:

15 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF33 shows 90.9% identity over a 143aa overlap with an ORF (ORF33a) from strain A of *N. meningitidis*:

20 orf33.pep LFLRVKVGRRFSSPATWFRXKDPVNQAVLR  
 orf33a LMDNQGLNFFLVLAVGXGMNTLMLAVWLAMLFLRVKVGRRFSSPATWFRGKDPVNQAVLR  
 90 100 110 120 130 140  
 25 orf33.pep LYXDEWRXTSVRWKIXATSHSLWLCTLLGMLVSVLLLLLVLRQYTFNWESTLLSNAASVRA  
 orf33a LYADEWRXPSVRWKIGATSHSLWLCTLLGMLVSVLLLLLVLRQYTFNWESTLLGDSSSVRL  
 150 160 170 180 190 200  
 30 orf33.pep VEMLAWLPSKLGFPVPDARAVIEGRNLGNIAADARAWSGLLVXSIACXGILPRL  
 orf33a VEMLAWLPAKLGFPVPDARAVIEGRNLGNIAADARAWSGLLVGSIACYGILPRLAWAVCK  
 210 220 230 240 250 260  
 35 orf33a ILXXTSENGLDLEKXXXXXIRRWQNKITDADTRRETQSAVSPKIVLNDAPKWAVMLETE  
 270 280 290 300 310 320

The complete length ORF33a nucleotide sequence <SEQ ID 201> is:

40 1 ATGTGAATC CATCCGAAA ACTGGTTGAG CTGGTCCGTA TTTTGAAGA  
 51 AGGCGGCTTT ATTTTCAGCG GCGATCCCGT GCAGGCGACG GAGGCTTTGC  
 101 GCCGCGTGGA CGGCAGTACG GAGGAAAAAA TCATCCGTCG GGCGAAGATG  
 151 ATCGACAGGA ACCGTATGCT GCGGAGACG TTGGAACGTG TCGGTGCGGG  
 201 GTCGTTCTGG TTGTGGGTGG CGGCGGCGAC GTTGCGTTT NTTACCGNTT  
 251 TTTTCAATTAC TTATCTTCTA ATGACAATC AGGGTCTGAA TTTCTTTTGT  
 45 301 GTTTTGGCGG GCGTGTGCGG CATGAATACG CTGATGCTGG CAGTATGGTT  
 351 GGCAATGTTG TTCTGCGCGG TGAAGTGGG GCGTTTTTTC AGCAGTCCGG  
 401 CGACGTGGTT TCGGGGCAAA GACCCTGTCA ATCAGGCGGT GTGCGGCTG  
 451 TATGCGGACG AGTGGCGGCG ACCTTCGGTA CGTTGGAATA TAGGCGCAAC  
 501 TCAGCACAGC CTGTGGCTCT GCACGCTGCT CGGAATGCTG GTGTCGTAT  
 551 TGTGCTGCT TTTGGTGGCG CAATATACGT TCAACTGGGA AAGCACGCTG  
 601 TTGGGCGATT CGTCTTCGGT ACGCTGGTG GAAATGTTGG CATGGCTGCC  
 651 TGCGAAACTG GGTTCCTCCG TGCTGATGC GCGGCGGTC ATCGAAGGTC  
 701 GTCTGAACGG CAATATTGCC GATGCGCGGG CTGGTGGG GCTGCTGGTC  
 751 GGCAGTATCG CCGCTACGG CATCTGCCG CGCTCTTGG CTGGGCGGT  
 55 801 ATGCAAAATC CTNTGNAAA CAAGCGAAAA CGGCTTGAT TTGGAAGAGC  
 851 NNNNNNTCN NNCNTCATC CGCGCTGGC AGAACAAAA CACCGATGCG  
 901 GATACGCGTC GGGAAACCGT GTCCGCCGTT TCGCCGAAAA TCGTCTTGAA  
 951 CGATGCGCGG AAATGGGCGG TCATGCTGGA GACCGAATGG CAGGACGGCG  
 1001 AATGGTTCGA GGGCAGGCTG GCGCAGGAAT GGCTGGATAA GGGCGTTGCC  
 60 1051 GCCAATCGGG AACAGGTTGC CGCGCTGGAG ACAGAGCTGA AGCAGAAAC



1101 GGC GCAACTG CTTATCGGCG TGCGCGCCCA AACTGTGCCC GACCGCGGCG  
1151 TGTTCGGCA GATCGTCCGA CTTTCGGAAG CGGCGCAGGG CGGCGCGGTG  
1201 GTGCANCTTT TGGCGGAACA GGGGCTTTCA GACGACCTTT CGGAAAAGCT  
1251 GGAACATTGG CGTAACGCGC TGACCGAATG CGGCGCGGCG TGGCTGGAAC  
1301 CCGACAGAGC GGC G CAGGAA GGCCGTCTGA AAACCAACGA CCGCACTTGA

This encodes a protein having amino acid sequence <SEQ ID 202>:

1 MLNPSRK LVE LVRILEEGGF IFSGDPVQAT EALRRVDGST EEKIIRRAKM  
51 IDRNRLRET LERVAGSEFW LWVAAATFAF XT XFSVTYLL MDNQGLNFFL  
101 VLAGVXGMNT LMLAVWLAML FLRVKVG RFF SSPATWFRGK DPVNQAVLRL  
151 YADEWRXPSV RWKIGATSHS LWLCTLLGML VSVLLLLLVR QYTFNWESTL  
201 LGDSSSVRLV EMLAWLPAKL GFPVPDARAV IEGR LNGNIA DARAWSGLLV  
251 GSIACYGILP RLLAWAVCKI LXXTSENGLD LEKXXXXXXI RRWQNKITDA  
301 DTRRET VSAV SPKIVLNDAP KWAVMLETEW QDGEWFEGRL AQEWLDKGVA  
351 ANREQVALE TELKQKPAQL LIGVRAQTVP DRGVL RQIVR LSEAAQGGAV  
401 VXL LAEQGLS DDLSEKLEHW RNALTECGAA WLEPDRAAQE GR LKTNDRT\*

ORF33a and ORF33-1 show 94.1% identity in 444 aa overlap:

20 orf33a.pep 10 20 30 40 50 60  
MLNPSRK LVE LVRILEEGGF IFSGDPVQATEALRRVDGST EEKIIRRAKMIDRNRLRET  
orf33-1 10 20 30 40 50 60  
MLNPSRK LVE LVRILEEGGF IFSGDPVQATEALRRVDGST EEKIIRRAEMIDRNRLRET

25 orf33a.pep 70 80 90 100 110 120  
LERVRAGSFWLWVAAATFAFXTXFSVTYLLMDNQGLNFFLVLAGVXGMNTLMLAVWLAML  
orf33-1 70 80 90 100 110 120  
LERVRAGSFWLWVAAATFAFFTGF SVTYLLMDNQGLNFFLVLAGVLGMNTLMLAVWLAML

30 orf33a.pep 130 140 150 160 170 180  
FLRVKVG RFF SSPATWFRGKDPVNQAVLRLYADEWRXPSVRWKIGATSHSLWLCTLLGML  
orf33-1 130 140 150 160 170 180  
FLRVKVG RFF SSPATWFRGKDPVNQAVLRLYADEWRQPSVRWKIGATSHSLWLCTLLGML

35 orf33a.pep 190 200 210 220 230 240  
VSVLLLLLVRQYTFNWESTLLGDSSSVRLV EMLAWLPAKL GFPVPDARAV IEGR LNGNIA  
orf33-1 190 200 210 220 230 240  
VSVLLLLLVRQYTFNWESTLLSNAASVRAV EMLAWLPSKL GFPVPDARAV IEGR LNGNIA

40 orf33a.pep 250 260 270 280 290 300  
DARAWSGLLVGSIACYGILPRL LAWAVCKI LXXTSENGLDLEKXXXXXXI RRWQNKITDA  
orf33-1 250 260 270 280 290 300  
DARAWSGLLVGSIACYGILPRL LAWVCKI LLKTS ENGLDLEK PYYQAVIRRWQNKITDA

45 orf33a.pep 310 320 330 340 350 360  
DTRRET VSAV SPKIVLNDAP KWAVMLETEW QDGEWFEGRLAQEWLDKGVAANREQVALE  
orf33-1 310 320 330 340 350 360  
DTRRET VSAV SPKII LNDAP KWAVMLETEW QDGEWFEGRLAQEWLDKGVATNREQVALE

50 orf33a.pep 370 380 390 400 410 420  
TELKQKPAQL LIGVRAQTVP DRGVL RQIVRLSEAAQGGAVVXLLAEQGLSDDLSEKLEHW  
orf33-1 370 380 390 400 410 420  
TELKQKPAQL LIGVRAQTVP DRGVL RQIVRLSEAAQGGAVVQLLAEQGLSDDLSEKLEHW

55 orf33a.pep 430 440 450  
RNALTECGAAWLEPDRAAQEGR LKTNDRTX  
orf33-1 430 440  
RNALAE CGAAWLEPDRAAQEGR LKQDX

Homology with a predicted ORF from *N.gonorrhoeae*

ORF33 shows 91.6% identity over a 143aa overlap with a predicted ORF (ORF33.ng) from *N.gonorrhoeae*:

5	orf33.pep	LFLRVKVGRRFFSSPATWFRXKDPVNQAVLR	30
	orf33ng	LMDNQGLNFFLVLAVGLGMNTLMLAVWLATLFLRVKVGRRFFSSPATWFRGKGPVNQAVLR	100
	orf33.pep	LYXDEWRXTSVRWKIXATSHSLWLCTLLGMLVSVLLLLLVLRQYTFNWESTLLSNAASVRA	90
10	orf33ng	LYADQWRQPSVRWKIGATAHSWLCTLLGMLVSVLLLLLVLRQYTFNWESTLLSNAASVRA	160
	orf33.pep	VEMLAWLPSKLGFPVPDARSVIEGRNLNGNIADARAWSGLLVXSIACTXGILPRL	143
	orf33ng	VEMLAWLPSKLGFPVPDARAVIEGRNLNGNIADARAWSGLLVGSIVCYGILPRLLAUVVCK	220

- 15 An ORF33ng nucleotide sequence <SEQ ID 203> was predicted to encode a protein having amino acid sequence <SEQ ID 204>:

```

1  MIDRDRMLRD  TLERVRAGSF  WLWVVVASMM  FTAGFSGTYL  LMDNQGLNFF
51  LVLAVGLGMN  TMLAVWLAT  LFLRVKVGRR  FSSPATWFRG  KGPVNQAVLR
101 LYADQWRQPS  VRWKIGATAH  SLWLCTLLGM  LSVLLLLLV  RQYTFNWEST
151 LLSNAASVRA  VEMLAWLPSK  LGFPVPDARA  VIEGRNLNGN  ADARAWSGLL
201 VGSIVCYGIL  PRLLAUVVCK  ILKTSSENG  DLEKTYQAV  IRRWQNKITD
251 ADTRRETUSA  VSPKIVLND  PKWALMLETE  WQDQWFEGR  LAQEWLDKGV
301 AANREQVAAL  ETELKQKPAQ  LLIGVRAQTV  PDRGVLRQIV  RLSEAAQGGV
351 VVQLLAEQGL  SDDLSEKLEH  WRNALTECGA  AWLEPDRVAQ  EGRKDKDQ*

```

- 25 Further sequence analysis revealed the following DNA sequence <SEQ ID 205>:

```

1  ATGTTGaatC  CATCCCgaAA  ACTGgttgag  ctGgTCCgtA  Ttttgaataa
51  aggggggtTTT  attttcagcg  gcgatcctgt  gcaggcgacg  gaggctttgc
101 gccgcgtgga  cggcAGTACG  GAggAaaaaa  tcttccgtcg  GGCGGAGAtg
151 atcgACAGGg  accgtatgtt  gcgggACaCg  TtggaaacGTG  TGCGTGCGgg
201 gtcgtTctgG  TTATGGGTGG  TggtggCatC  gATGATGttt  aCGCCGGAT
251 TTTCAGgcac  ttatCttCTG  ATGGAcaatC  AGGGGctGAA  TtTCTTTTtA
301 GTTTtgcgG  GAGTgttgG  CATGaatacG  ctgATGCTGG  CAGTATGGtt
351 gGCAACGTTG  TTCTGCGCG  TGAAAGTGGG  ACGGTTTTTC  AGCAGTCCGG
401 CGACGTGgTT  TCGGGGCAAA  GGCCCTGTAA  ATCAGGCGGT  GTTGGCGCTG
451 TATGCGGACC  AGTGGCGGCA  ACCTTCGGTA  CGATGGAAAA  TAGGCGCAAC
501 GGCGCACAGC  TTGTGCTCT  GCACGTGCT  CGGAATGCTG  GTGTCTGCTAT
551 TGCTGCTGCT  TTTGGTGGCG  CAATATACGT  TCAACTGGGA  AAGCAGCTG
601 TTGAGCAATG  CCGCTTCGGT  ACGCGCGGTG  GAAATGTTGG  CATGGCTGCC
651 GTCGAAACTC  GGTTCCTCT  TCCCCGATGC  GCGGCGCGTC  ATCGAAGGTC
701 GTCTGAACGG  CAATATTGCC  GATGCGCGGG  CTGGTCTGG  GCTGCTGCTC
751 GGCAGTATCG  TCTGCTACGG  CATCCTGCCG  CGCCTCTTGG  CTTGGGTAGT
801 GTGTAAATC  CTTTGAATA  CAAGCGAAAA  CGGattgGAT  TTGGAAAAAA
851 CCTATTATCA  GGCGGTCATC  CGCCGCTGGC  AGAACAAAAT  CACCGATGCG
901 GATACGCGTC  GGGAAACCGT  GTCCGCCGTT  TCGCcgaAAA  TCGTCTTGAA
951 CGATGCGCGC  AAATGGGCGC  TCATGCTGGA  GACCGAGTGG  CAGGACGGCC
1001 AATGGTTCGA  GGGCAGGCTG  GCGCAGGAAT  GGCTGGATAA  GGGCGTTGCC
1051 GCCAATCGGG  AACAGGTGTC  CGCGCTGGAG  ACAGAGCTGA  AGCAGAAACC
1101 GGCGCAACTG  CTTATCGGCG  TACGCGCCCA  AACTGTGCCG  GACCGGGGCG
1151 TGCTGCGGCA  GATTGTGCGG  CTTTCGGAAG  CGGCGCAGGG  CGGCGCGGTG
1201 GTGACGTTT  TGGCGGAACA  GGGGCTTTCA  GACGACCTTT  CGGAAAAGCT
1251 GGAACATTGG  CGTAACGCGC  TGACCGAATG  CGGCGCGGCG  TGGCTTGAGC
1301 CTGACAGGGT  GGCGCAGGAA  GGCGGTTTGA  AAGACCAATA  A

```

This encodes a protein having amino acid sequence <SEQ ID 206; ORF33ng-1>:

```

1  MLNPSRKLVE  LVRILNKGFF  IFSGDPVQAT  EALRRVDGST  EEKIFRAEM
51  IDRDRMLRDT  LERVRAGSFW  LWVVVASMME  TAGFSGTYLL  MDNQGLNFFL
101 VLAGVLGMNT  LMLAVWLATL  FLRVKVGRRFF  SSPATWFRGK  GPNQAVLRL
151 YADQWRQPSV  RWKIGATAHS  LWLCTLLGML  VSVLLLLLV  QYTFNWESTL
201 LSNAASVRVA  EMLAWLPSKL  GFPVPDARAV  IEGRNLNGNIA  DARAWSGLLV

```

251 GSIVCYGILP RLLAWVCKI LLKTSEGLD LEKTTYQAVI RRWQNKITDA  
 301 DTRRETSAV SPKIVLNDAP KWALMLETEW QDGQWFEGR LAQEWLDKGVA  
 351 ANREQVALE TELKQKPAQL LIGVRAQTVP DRGVLRQIVR LSEAAQGGAV  
 401 VQLLAEQGLS DDLSEKLEHW RNALTECGAA WLEPDRVAQE GRLKDQ\*

5 ORF33ng-1 and ORF33-1 show 94.6% identity in 446 aa overlap:

		10	20	30	40	50	60
	orf33-1.pep	MLNPSRKLV	LRILDEGGF	IFSGDPVQ	ATEALRRVD	GSTEEKIIR	RAEMIDRNR
	orf33ng-1	MLNPSRKLV	LRILNKGGF	IFSGDPVQ	ATEALRRVD	GSTEEKIFR	RAEMIDRDR
10		10	20	30	40	50	60
	orf33-1.pep	LERVRAGSF	WLWVVAAT	FAFFTGF	SVTYLLMD	NQGLNFFL	VLAGVLGMN
	orf33ng-1	LERVRAGSF	WLWVVASMM	FTAGFSGT	YLLMDNQ	GLNFFLV	LAGVLGMN
15		70	80	90	100	110	120
	orf33-1.pep	LERVRAGSF	WLWVVAAT	FAFFTGF	SVTYLLMD	NQGLNFFL	VLAGVLGMN
	orf33ng-1	LERVRAGSF	WLWVVASMM	FTAGFSGT	YLLMDNQ	GLNFFLV	LAGVLGMN
20		130	140	150	160	170	180
	orf33-1.pep	FLRVKVG	RFFSSPAT	WFRGKDP	VNQAVLR	LYADEWR	QPSVRWK
	orf33ng-1	FLRVKVG	RFFSSPAT	WFRGKDP	VNQAVLR	LYADEWR	QPSVRWK
25		190	200	210	220	230	240
	orf33-1.pep	VSVLLLL	LVRYQY	TFNWEST	LLSNAAS	VRVEM	LAWLPSKL
	orf33ng-1	VSVLLLL	LVRYQY	TFNWEST	LLSNAAS	VRVEM	LAWLPSKL
30		250	260	270	280	290	300
	orf33-1.pep	DARAWSG	LLVGSIA	CGILPRL	LAWVVKI	LLKTSEN	GLDLEKPY
	orf33ng-1	DARAWSG	LLVGSIA	CGILPRL	LAWVVKI	LLKTSEN	GLDLEKPY
35		310	320	330	340	350	360
	orf33-1.pep	DTRRET	SAVSPKI	ILNDAP	KWAVMLE	TEWQDGE	WFEGRLA
	orf33ng-1	DTRRET	SAVSPKI	ILNDAP	KWAVMLE	TEWQDGE	WFEGRLA
40		370	380	390	400	410	420
	orf33-1.pep	TELKQK	PAQLLIG	VRAQTVP	DRGVLR	QIVRLSE	AAQGGAV
	orf33ng-1	TELKQK	PAQLLIG	VRAQTVP	DRGVLR	QIVRLSE	AAQGGAV
45		430	440				
	orf33-1.pep	RNALAEC	GAAWLEP	DRAAQE	GRLKDQX		
	orf33ng-1	RNALAEC	GAAWLEP	DRAAQE	GRLKDQX		
50		430	440				

Based on the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 25

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 207>:

1 ..CAGAAGAGTT TGTCGAGAAT TTCTTTATGG GGTGGGCG GCGTGT

51 CGGGGTGTCC GGTCTGGTAT GGTTCCTTT GGGCGTTTCT TT.GAGTGCG  
 101 CCTGTTTTTC GGGTGTTCCT TTTCGGGGT CGGGACGGGG GACGTTTGTG  
 151 GGCAGTACGG GGGTTTCTTT GAGTGTGTT TCAGCTTGTG TTCC.GGCGT  
 201 CGTCCGGCTG CCTGTCGGT TGAGCTGTG CGGCAGGTG CG..GTTTGA  
 251 CCCGGTTTTT CTTGGGTGCG GCAGGGGACG TCATTCTCCT GCCGCTTTCG  
 301 TCTGTGCCGT CCGCTGTGC GGGTTCGGAT GAGGCGGCGT GGTGGTGTTC  
 351 GGGTTGGGCG GCATCTTGT CCGACTACGC CGTTTGGCAG CCAGAATTCTG  
 401 GTTTCGCGG GGCTGTCCGT GTGTTGCGGT TCGGCTTGAA GGGTTTTGTC  
 451 GTCC..

10 This corresponds to the amino acid sequence <SEQ ID 208; ORF34>:

1 ..QKSLSRISLW GLGGVFFGVS GLVWFSLVGS XECACFSGVS FRGSGRGTFV  
 51 GSTGVSLSVF SACVXGVVRL PVGLSCVGR LXXLTRFFLGA AGDVILLPLS  
 101 SVPSGCAGSD EAAWCSGWA ASCPTTFFGS QNSVSRGLSV CCGSA\*RVLS  
 151 S..

15 Further work revealed the complete nucleotide sequence <SEQ ID 209>:

1 ATGATGATGC CGTTCATAAT GCTTCCTTGG ATTGCGGGTG TGCCTGCCGT  
 51 GCCGGGTCAG AATAGGTTGT CCAGAATTC TTTATGGGGT TTGGGCGGCG  
 101 TGTTTTTCGG GGTGTCCGGT TTGGTATGGT TTTCTTTGGG CGTTTCTTTG  
 151 GGCTGCGCCT GTTTTTCCGG TGTTCTTTT CGGGGTTCCG GACGGGGGAC  
 201 GTTTGTGGG AGTACGGGGG TTTCTTTGAG TGTGTTTCA GCTTGTGTTT  
 251 CGGCTCGTC CGGCTGCCTG TCGTTTGGG CTGTGTCGGC AGGTTGCGGT  
 301 TTGACCGGT TTTCTTTGGG TGCGGCAGGG GACGGCAGTC CGCTGCCGCT  
 351 TTCGTCTGT CCGTCCGGCT GTGCGGGTTC GGATGAGGCG GCGTGGTGGT  
 401 GTTCGGGTTG GCGGCGATCT TGTCGACTA CGCCGTTTGG CAGCCAGAAT  
 451 TCGGTTTCG GGGGGCTGTC GGTGTGTTGC GGTTCGGCTT GAAGGGTTTT  
 501 GTCGCCGTTT GGGTTGAATG TGCTGACGAT GCCTATTGCC AATGCGCCGA  
 551 TGGCGCGGAT ACAGATGAGC AATACGGCGC GTATCAGGAG TTGGGGGTC  
 601 AGCCTGAAGG GTTTGTTCCG TTTTCTTGGC ATTTTGATTG TGCTTTTGGG  
 651 GTGTCGGGCA ATGCCGTCTG AAGGCGGTTT AGACGGCATT GCCGAGTCAG  
 701 CGTTGGACGT AGTTTGGTA GAGGGTGATG ACTTTTGTG CCGCGACGGT  
 751 GGTGCTGACT TTTGGGTAA TCTGCGCCTG TTCTTCGGGG GTGAGGATGC  
 801 CCATAACGTA GGTACGTTG CCGTAGGTAA CGATTTTAC GCGCGCCTGT  
 851 GTGGCGGGG TGATGCCCAA CAGCGTGCGC CGGACTTTGG ATGTGTTCCA  
 901 AGTGTGCGG GCGATGTGCG CGGCAGTGCG CGGCAGGAG GCGACGGTAA  
 951 TATAGTTGTA CACGCCTTCG GCGGCCTGTT CGGAACGTGC AATCTGACCG  
 1001 ACGAACTGTT TTTCGCCTTC GGTGGCGACT TGTCGAGCA GCAGCAGGTG  
 1051 GCGGTTGTAG CCGACGACGG AGATTGCGG CGTGTAGCCT TTGGTTTGGT  
 1101 TGTTTTGGCG CAGATAGGAA CGGGCGGTGG TTTTCGATAC CAACGCCATA  
 1151 ACGTTGTCGT CGGTTTGGCG GCCGGTGGT CGGCGGTCGA CGGCGGATTT  
 1201 CGCGCCGACG GCGGCGCTTC CGATTACTGC GCTGACGCG CCGCTAAGGG  
 1251 CAAGGCTGAA AATGGCGGCA ATCAGGGTGC GGACGGTGTG CGGTTTGGGT  
 1301 TTCATCGGGT CATTCTTTC TTGGGCGTTT CAGACGGCAT TGCTTTGCGC  
 1351 CATGCCGTCT GA

This corresponds to the amino acid sequence <SEQ ID 210; ORF34-1>:

45 1 MMMPFIMLPW IAGVPAVPGQ NRLSRISLWG LGGVFFGVSG LVWFSLVGSL  
 51 GCACFSGVSF RSGRGTFVG STGVSLSVFS ACVPASSGCL SV\*AVSAGCG  
 101 LTRFFLGAAG DGSPLPLSSV PSGCAGSDEA AWWCSGWAAS CPTTFFGSQN  
 151 SVSRGLSVCC GSA\*RVLSPF GLNVLTMPA NAPMAAIQMS NTARIRSLGV  
 201 SLKGLFGFFA ILIVLLGCRA MPSEGGSDGI AESALDVVLV EGDDFLYADG  
 50 251 GADFLGNLRL FFGGEDAHNV GYVAVGNDFD ARLCGGADAQ QRGADFGCVP  
 301 SVAGDVAGSA RQGGDGNIVV HAFGGLFGTC NLTDLEFFAF GGDLEQQQV  
 351 AVVADDGDLG RFAFGLVLA QIGTGGGFDI QRHNVVVGLR AGGSVDGGF  
 401 RADGGASDYC ADAAAKGRAE NGGNQAGADV RFGFHRVLPF LGVSDGIALR  
 451 HAV\*

55 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF34 shows 73.3% identity over a 161aa overlap with an ORF (ORF34a) from strain A of *N. meningitidis*:

orf34.pep  
 orf34a  
 orf34.pep  
 orf34a  
 orf34.pep  
 orf34a  
 orf34.pep  
 orf34a  
 orf34.pep  
 orf34a  
 orf34.pep  
 orf34a  
 orf34.pep  
 orf34a

10 20 30 40 50 60  
 40 50 60 70 80 90  
 100 110 120 130 140 150  
 120 130 140 150 160 170  
 180 190 200 210 220 230

QKSLSRISLWGLGGVFFGVSGLVWFSLVGSXE-----CAC  
 MMXPXIMLPWIAGVPAVPGQKRLSRXSLWGLGGXFFGVSGLVWFSLVGSXSXSLGVXGXCAC  
 FSGVSFRGSGRGT FVGSTGVSLSVFSACVXGVVRLPVGLSCVGRLLX-----LTRFFFLGA  
 FSGVSFRGSGRGT FVGSTGVSLSVFSACA-----PASSGCLSVKAVSAGCGLTRXFXGA  
 AGDVILLPLSSVPSGCAGSDEAAWWC SGWAASCPTT PFGSQNSVSRGLSVCCGSAXRVLS  
 AGDGSPLPLSSVPSGCAGADEEAXXC SGWAASCPTT PFGSQNSVSRGLSVCCGSVVRVLS  
 S  
 PFGXNVLTMPIANAPMAVIQMSNTARISRLGVSILKGLXFFAILIVLLGCRAMPSEGGSD

The complete length ORF34a nucleotide sequence <SEQ ID 211> is:

1 ATGATGATNC CGTTNATAAT GCTTCCTTGG ATTGCGGGTG TGCCTGCCGT  
 51 GCCGGGTCAG AAGAGGTTGT CGAGAANTTC TTTATGGGGT TTAGGCGGCN  
 101 TGTTTTTCGG GGTGTCCGGT TTGGTATGGT TTTCTTTGGG CGTTTCTNTT  
 151 TCTTTGGGTG TTTCTNTGGG CTGTGCCTGT TTTTCGGGTG TTTCTTTTCG  
 201 GGGTTCGGGA CGGGGGACGT TTGTGGGCAG TACNGGGGT TCTTTAGTG  
 251 TGTTTTCAGC TTGTGCTCCG GCGTCGTCCG GCTGCCTGTC GGTTNAGCT  
 301 GTGTCCGCAG GTTGCCTTT GACCCGGNTT TTCTTNGGTG CGGCAGGGGA  
 351 CGGCAGTCCG CTGCCGCTTT CGTCTGTGCC GTCCGGCTGT GCGGGTGCGG  
 401 ATGAGGAGGC GTNGTNGTGT TCGGGTTGGG CGGCATCTTG TCCGACTACG  
 451 CCGTTTGGCA GCCAGAATTC GGTTCGCGG GGGCTGTCCG TGTGTTGCGG  
 501 TTCGGTNTGG AGGGTTTTGT CNCCGTTCCG GTNGAATGTG CTGACGATGC  
 551 CTATTGCCAA TCGCCCGATG GCGGTGATAC AGATGAGCAA TACGGCGCGT  
 601 ATCAGGAGTT TGGGGGTCAG CCTGAAGGGT TTGTTCTNGT TTTTGGCCAT  
 651 TTTGATGTGT CTTTGGGGT GTCCGGCAAT GCCGTCTGAA GCGGTTTCAG  
 701 ACGGCATTGC CGAGTCAGCG TTGGACGTAG TTTNGGTAGA GGGTGATGAC  
 751 TTTTGTACG CCGACCGTGG TGCTGACTTT TTGGGTAATC TCGCCTGTT  
 801 CTTCCGGGGT GAGGATGCCC ATAACGTAGG TTACGTTGCC GTAGGTAACG  
 851 ATTTTACGCG GCGCCTGTGT GCGGGGGCTG ATGCCCAACA GCGTGGCGCG  
 901 GACTTTGGAT GTGTCCAAG TGTCGCGCGC GATGTCGCGC GCAGTGCGCG  
 951 GCAGGGAGGC GACGGTAATG TANTGTACA CGCCTTCGCG GGCCTGTTTCG  
 1001 GAACGTGCAA TCTGACCGAC GAACTGTTTC TCGCCTTCGG TGGCGACTTG  
 1051 TCCGAGCAGC AGCAGGTGGC GGTGTAGCC GACAACGGAG ATTTGGGGCG  
 1101 TGTANCCTTT GGTTTGGTTG TTTTGGCGCA GATAGGAGCG GCGGTTGGTT  
 1151 TCGATACGCA GCGCCATTAC GTTGTCTGCG GTTNGCGCGC CGGTGGTTTCG  
 1201 GCGGTCGACG GCGGATTTCG CGCCGACCGC CGCGCCGCGC ACGACTGCGC  
 1251 TGACGAGGCC GCCGAGGGCA AGGCTGAGGA CGGCGGCAGT CAGGGTGCGG  
 1301 ACGGTGTGCG GTTTGGGTTT CATCGGGTGC TTCTTTTCTT GGGCGTTTCA  
 1351 GACGGCATTG CTTTGCGCCA TGCCGTCTGA

This encodes a protein having amino acid sequence <SEQ ID 212>:

1 MMXPXIMLPW IAGVPAVPGQ KRLSRXSLWG LGGXFFGVSG LVWFSLVGSX  
 51 SLGVXGXCAC FSGVSFRGSG RGT FVGSTGV SLVFSACAP ASSGCLSVXA  
 101 VSAGCGLTRX FXGAAGDGSP LPLSSVPSGC AGADEEAXXC SGWAASCPTT  
 151 PFGSQNSVSR GLSVCCGSVW RVLSPFGXNV LTMPIANAPM AVIQMSNTAR  
 201 IRS LGVSLKG LFXFFAILIV LLGCRAMPSE GSDGIAESA LDVXXVEGDD  
 251 FLYADGGADF LGNLRLLFFGG EDHNVGYVA VGNDFDARLC GGADAQQRGA  
 301 DFGCVPSVAG DVAGSARQGG DGNVXVHAFG GLFGTCNLTD ELFLAFGGDL  
 351 SEQQQVAVVA DNGDLGRVXF GLVVLAQIGA GGGFDTQRHY VVVGXRAGGS  
 401 AVDGGFRADR RAADDCAADAA AEGKAEDGGS QGADGVRFGF HRVLPFLGV  
 451 DGIALRHAV\*

ORF34a and ORF34-1 show 91.3% identity in 459 aa overlap:

		10	20	30	40	50	60
	orf34a.pep	MMXPXIMLPWIAGVPAVPGQKRLSRXSLWGLGGXFFGVSGLVWVFSGLVGSXSXSGVSGXCAC					
5	orf34-1	MMMPFIMLPWIAGVPAVPGQKRLSRISLWGLGGVFFGVSGLVWVFSGLVGSXSL-----GCAC					
		10	20	30	40	50	
	orf34a.pep	FSGVSFRSGRGTFVGSTGVSLSVFSACAPASSGCLSVXAVSAGCGLTRFXGAAGDGSP					
10	orf34-1	FSGVSFRSGRGTFVGSTGVSLSVFSACVPASSGCLSVXAVSAGCGLTRFFLGAAGDGSP					
		60	70	80	90	100	110
	orf34a.pep	LPLSSVPSGCAGADEEAXXCSGWAASCPTTPFGSQNSVSRGLSVCCGSVWRVLSPFGXNV					
15	orf34-1	LPLSSVPSGCAGSDEAAWCSGWAASCPTTPFGSQNSVSRGLSVCCGSAXRVLSPFGLNV					
		120	130	140	150	160	170
	orf34a.pep	LTMPIANAPMAVIQMSNTARIRSLGVSLKGLFEXFFAILIVLLGCRAMPSEGGSDGIAESA					
20	orf34-1	LTMPIANAPMAAIQMSNTARIRSLGVSLKGLFGFFAILIVLLGCRAMPSEGGSDGIAESA					
		180	190	200	210	220	230
	orf34a.pep	LDVVXVEGDDFLYADGGADFLGNLRLFFGGEDAHNVGYVAVGNDFDARLCGGADAQQRGA					
25	orf34-1	LDVVLVEGDDFLYADGGADFLGNLRLFFGGEDAHNVGYVAVGNDFDARLCGGADAQQRGA					
		240	250	260	270	280	290
	orf34a.pep	DFGCVPSVAGDVAGSARQGGDGNVXVHAFGGLFGTCNLTDLFLAFGGDLSEQQQVAVVA					
30	orf34-1	DFGCVPSVAGDVAGSARQGGDGNIVVHAFGGLFGTCNLTDLFFAFGGDLSEQQQVAVVA					
		300	310	320	330	340	350
	orf34a.pep	DNGDLGRVXFGLVVLQAIGAGGGFDTQRHYVVVGXRRAGGSAVDGGFRADRRADDCAADAA					
35	orf34-1	DDGDLGRVAFGLVVLQAIGTGGGFDTQRHNVVVGLRAGGSAVDGGFRADGGASDYCADAA					
		360	370	380	390	400	410
	orf34a.pep	AEGKAEDGGSQGADGVRFGFHRVLPFLGVSDGIALRHAVX					
40	orf34-1	AKGKAENGNGQADGVRFGFHRVLPFLGVSDGIALRHAVX					
		420	430	440	450		

# Homology with a predicted ORF from *N. gonorrhoeae*

50 ORF34 shows 77.6% identity over a 161aa overlap with a predicted ORF (ORF34.ng) from *N. gonorrhoeae*:

	orf34.pep	QKSLSRISLWGLGGVFFGVSGLVWVFSGLVGSXE-----CAC	35
	orf34ng	MMMPFIMLPWIAGVPAVPGQKRLSRISLWGLAGVFFGVSGLVWVFSGLVGSFSLGVSLGCAC	60
55	orf34.pep	FSGVSFRSGRGTFVGSTGVSLSVFSACVXGVRLPVGLSCV-----GRLXXLTRFFLGA	90
	orf34ng	FSGVSFRSGGWGAFVGSTGVSLSVFSACVP-----VPVNESAARAASEGR--GLTRFFLGA	114
60	orf34.pep	AGDVILLPLSSVPSGCAGSDEAAWCSGWAASCPTTPFGSQNSVSRGLSVCCGSAXRVLS	150
	orf34ng	AGDGSPLPLSSVPSGCAGSDEAAWCSGWAASCPTAPFGSQNSVSRGLSVCCGSVWRVLS	174
65	orf34.pep	S	175
	orf34ng	PFGLNVLTMTANAPMAVIQMSNTARIRSLGVSLKGLFGFFAILIVLLGCRAMPSEGGSD	234

The complete length ORF34ng nucleotide sequence <SEQ ID 213> is:

5  
10  
15  
20  
25  
30  
35  
40

```

1  ATGATGATGC CGTTCATAAT GCTTCCTTGG ATTGCGGGTG TGCCTGCCGT
51  GCCGGGTCAA AAGAGGTGTG CGAGAATCTC TTTATGGGGT TTGGCCGGCG
101 TGTTTTTCGG GGTGTCCGGT TTGGTATGGT TTTCTTTGGG CGTTTCTTTT
151 TCTTTGGGTG TTTCTTTGGG CTGCGCCTGT TTTTCGGGTG TTTCTTTTCG
201 GGGTTCGGGA TGGGGGGCGT TTGTGGGCAG TACGGGGGTT TCTTTGAGTG
251 TGTTTTTCAGC TTGTGTTCGG GTGCCGGTTA ACGAATCGGC TGCCCGGGCC
301 GCATCCGAAG GGCGCGGTTT gACCCGGTTT TTCTTGGGTG CGGCAGGGGA
351 CGGCAGTCCG CTGCCGCTTT CTCTGTGCC GTCCGGCTGT GCGGGTTCCG
401 ATGAGGCGGC GTGGTGGTGT TCGGGTTGGG CGGCATCTTG TCCGACGGCG
451 CCGTTTGGCA GCCAGAATTC GGTTTCGGCG GGGCTGTCTG TGTGTGCGG
501 TTCGGTTTGG AGGGTTTGTG CGCCGTTCCG GTTGAATGTG CTGACGATGC
551 TCTACTGCCAA TGGCCTCATG GCGGTGATAC AGATGAGCAA TACGGCGCGT
601 ATCAGGAGTT TGGGGGTGAG CCTGAAGGGT TTGTTCGGTT TTTTGTCCAT
651 TTTGATTGTG CTTTTGGGGT GTCGGGCAAT GCCGTCTGAA GCGGGTTCAG
701 ACGGCATTGC CGAGTTCAGC TTGGACGTAG TTTGGGTAGA GGGTAATGAC
751 TTTTGTACG CCGAcggTGG TGCTGACTTT TTGGGTAATC TCGCCTGTT
801 CTTCGGGGGT GAGGATGCCC ATAACGTAGG TTACATTGCC GTAGGTAATG
851 ATTTTGACGC GCGCCTGTGT AGCGGGGCTG ATGCCAGCA GcgtgGCGCG
901 GACTTTGGAC GTGTTCCAAG TGTGCGCGGC GATGTCGCC GCAGTGCAGC
951 GCAGGGAGGC GACGGTAATG TAGTTGTATA CGCCTTCGGC GGCCTGTTTCG
1001 GAACGTGCAA TCTGACCGAC GAACTGTTTT TCGCCTTCGG TGGCGACTTG
1051 TCCGAGCAGC AGCAGTGGC GGTGTAGCC GACGACGGAG ATTTGGGGCG
1101 TGTAGCCTTT GGTTTGGTTG TTTTGGCGCA GGTAGGAACG GCGGTGTTT
1151 TCGATACGCA ACGCCATAAC GTtgtCATCG GTTtgcgcgc CGGTGGTtgc
1201 gCGGTCGATG ACGGATTTTG CGCCGACGGC GGGCCCGCCG ACGACTGCGC
1251 TGAAGCAGCC GCCGAGGCA AGGCTGAGGA CGGCGGCAAT CAGGTTGCGG
1301 ACGGTGTGTG GTTTGGGTTT CATCGGGGAC TTCCTTTCTT GGGCGTTTCA
1351 GACGGCATTG CTTTGCGCCA TGCCGTCTGA

```

This encodes a protein having amino acid sequence <SEQ ID 214>:

35  
40

```

1  MMMPFIMLPW IAGVPAVPGQ KRLSRISLWG LAGVFFGVSG LVWFSLGVSF
51  SLGVSLGCAC FSGVSFRGSG WGA FVGSTGV SLSVFSACVP VPVNEAARA
101 ASEGRGLTRF FLGAAGDGSP LPLSSVPSGC AGSDEAAWVC SGWAASCPA
151 PFGSQNSVSR GLSVCCGSVW RVLSPFGLNV LTMPTANAPM AVIQMSNTAR
201 IRSLGVSLKG LFGFFAILIV LLGCRAMPSE GSGDGAIESA LDVVLVEGND
251 FLYADGGADF LGNLRLEFGG EDAHNVGYIA VGNDFDARLC SGADAQQRGA
301 DFGRVPSVAG DVAR SARQGG DGNVVVYAFG GLFGTCNLTD ELFFAFGGDL
351 SEQQQVAVVA DDGDLGRVAF GLVVLAQVGT GGGEDTQRHN VVIGLRAGGS
401 AVDDGFCADG GPADDCAEAA AEGKAEDGGN QGADGVWFGF HRGLPFLGV
451 DGIALRHAV*

```

ORF34ng and ORF34-1 show 90.0% identity in 459 aa overlap:

45  
50  
55  
60

```

      10      20      30      40      4      50
orf34-1.pep  MMMPFIMLPW IAGVPAVPGQNRLSRISLWGLGGVFFGVSGLVWFSLGVSLGVCAC
      10      20      30      40      50      60
orf34ng      MMMPFIMLPW IAGVPAVPGQKRLSRISLWGLAGVFFGVSGLVWFSLGVSLGVCAC

      60      70      80      90      100     110
orf34-1.pep  FSGVSFRGSGRGTFTVGSTGVSLSVFSACVPASSGCLSVXAVSAGCGLTRFFLGAAGDGSP
      70      80      90      100     110     120
orf34ng      FSGVSFRGSGWGA FVGSTGVSLSVFSACVPVPVNEAARAASEGRGLTRFFLGAAGDGSP

      120     130     140     150     160     170
orf34-1.pep  LPLSSVPSGCAGSDEAAWCSGWAASCPPTPFGSQNSVSRGLSVCCGSAXRVLSPFGLNV
      130     140     150     160     170     180
orf34ng      LPLSSVPSGCAGSDEAAWCSGWAASCPPTPFGSQNSVSRGLSVCCGSVVRVLSPFGLNV

      180     190     200     210     220     230
orf34-1.pep  LTMPIANAPMAAIQMSNTARIRSLGVSLKGLFGFFAILIVLLGCRAMPSEGGSDGIAESA
      190     200     210     220     230     240
orf34ng      LTMPTANAPMAVIQMSNTARIRSLGVSLKGLFGFFAILIVLLGCRAMPSEGGSDGIAESA

```

		240	250	260	270	280	290
5	orf34-1.pep	LDVV	LEGGDFLYADGGADFLGNLRLFFGGEDAHNVGYVAVGNDFDARLCGGADAQQRGA				
	orf34ng	LDVV	LEGGDFLYADGGADFLGNLRLFFGGEDAHNVGYIAVGNDFDARLCGGADAQQRGA				
		250	260	270	280	290	300
10	orf34-1.pep	DFGCV	PSVAGDVAGSARQGGDGNIVVHAFGGFLGTCNLTDELFFAFGGDLSEQQQVAVVA				
	orf34ng	DFGRV	PSVAGDVARSARQGGDGNVVVYAFGGFLGTCNLTDELFFAFGGDLSEQQQVAVVA				
		310	320	330	340	350	360
15	orf34-1.pep	DDGDL	GRVAFGLVVLAAQIGTGGGFDTRHNVVVLRAAGSAVDGGFRADGGASDYCADAA				
	orf34ng	DDGDL	GRVAFGLVVLAAQVGTGGGFDTRHNVVIGLRAAGSAVDDGFCADGGFADDCAEAA				
		370	380	390	400	410	420
20	orf34-1.pep	AKGKA	ENGGNQGADGVRFGFHRVLPFLGVSDGIALRHAVX				
	orf34ng	AEGKA	EDGGNQGADGVWFGFHRGLPFLGVSDGIALRHAVX				
		430	440	450	460		

- 25 Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 26

- 30 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 215>:

```

1  ATGAAACCT TCTTCAAAAC CCTTCCGCC GCCGCACTCG CGCTCATCCT
51 CGCCGCCTGC GGATT.CAAA AAGACAGCGC GCCCGCCGCA TCCGCTTCTG
101 CCGCCGCCGA CAACGGCGCG GCGTAAAAAA GAAATCGTCT TCGGCACGAC
151 CGTCGGCGAC TTCGGCGGATA TGGTCAAAGA ACAAATCCAA GCCGAGCTGG
35 201 AGAAAAAAGG CTACACCGTC AACTGGTTCG AGTTTACCGA CTATGTACGC
251 CCGAATCTGG CATTTGGCTGA GGGCGAGTTG

```

This corresponds to the amino acid sequence <SEQ ID 216; ORF4>:

```

1  MKTFFKTLA AALALILAAC G.QKDSAPAA SASAADNGA AKKEIVFGTT
51 VGDFGDMVKE QIQAELEKKG YTVKLVEFTD YVRPNLALAE GEL

```

- 40 Further sequence analysis revealed the complete nucleotide sequence <SEQ ID 217>:

```

1  ATGAAACCT TCTTCAAAAC CCTTCCGCC GCCGCACTCG CGCTCATCCT
51 CGCCGCCTGC GCGGTCAAA AAGACAGCGC GCCCGCCGCA TCCGCTTCTG
101 CCGCCGCCGA CAACGGCGCG GCGAAAAAAG AAATCGTCTT CGGCACGACC
151 GTCGGCGACT TCGGCGATAT GGTCAAAGAA CAAATCCAAG CCGAGCTGGA
45 201 GAAAAAAGGC TACACCGTCA AACTGGTTCG GTTACCAGAC TATGTACGCC
251 CGAATCTGGC ATTGGCTGAG GCGGAGTTGG ACATCAACGT CTCCAACAC
301 AAACCTATC TTGACGACTT CAAAAAGAA CACAATCTGG ACATCACCAG
351 AGTCTTCCAA GTGCCGACCG CGCCTTTGGG ACTGTACCCG GGCAAGCTGA
401 AATCGCTGGA AGAAGTCAA GACGGCAGCA CCGTATCCGC GCCCAACGAC
50 451 CCGTCCAAC TCGCCCGCGT CTTGGTGATG CTCGACGAAC TGGGTTGGAT
501 CAAACTCAA GACGGCATCA ATCCGTTGAC CGCATCCAAA GCGGACATCG
551 CCGAGAACCT GAAAAACATC AAAATCGTCG AGCTTGAAGC CGCGCAACTG
601 CCGCGTAGCC GCGCCGACGT GGATTTTGCC GTCGTCAACG GCAACTACGC
651 CATAAGCAGC GGCATGAAGC TGACCGAAGC CCTGTTCCAA GAACCGAGCT
701 TTGCCTATGT CAACTGGTCT GCCGTCAAAA CCGCCGACAA AGACAGCCAA
751 TGGCTTAAAG ACGTAACCGA GGCCTATAAC TCCGACGCGT TCAAAGCCTA

```



801 CGCGCACAAA CGCTTCGAGG GCTACAAATC CCCTGCCGCA TGAATGAAG  
851 GCGCAGCCAA ATAA

This corresponds to the amino acid sequence <SEQ ID 218; ORF4-1>:

5	1	MKTFFKTL	SA	AALALILA	AC	GGQKDSAPAA	SASAAAADNGA	AKKEIVFGTT
	51	VGDFGDMVKE	QIQA	ELEKKG		YTVKLVLEFTD	YVRPNLALAE	GELDINVFQH
	101	KPYLLDDFKFE	HNLDI	TEVFQ		VKTAPLGLGYP	GKLKSLEEVK	DGSTVSAFND
	151	PSNFARVLVM	LDELG	WIKLK		DGINPLTASK	ADTAENLKNI	KIVELEAAQL
	201	PRSRADVDA	VVNGN	YAISS		GMKLTEALFQ	EPSFAYVNW	AVKTAADKDSQ
	251	WLKDVTEFY	SDNFK	AYAHK		RMGEYKSPAA	WNEGAAK*	

10 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF4 shows 93.5% identity over a 93aa overlap with an ORF (ORF4a) from strain A of *N. meningitidis*:

15	orf4.pep	<u>MKTFFFKTLSAAALALILAA</u> CG-QKDSAPAASASAAADNGAAKKEIVFGTTVGDFGDMVKE	10 20 30 40 50 59
	orf4a	<u>MKTFFFKTLSAAALALILAA</u> CGGQKDSAPAASASAAADNGAAXKEIVFGTTVGDFGDMVKE	10 20 30 40 50 60
20	orf4.pep	QIQAELEKKGYTVKLVFTDYVRPNLALAEGEL	60 70 80 90
	orf4a	XIQPELEKKGYTVKLVEXTDYVRXNLALAEGELDINVXQHXXYLDXKKXHNLDITXVXQ	70 80 90 100 110 120
25	orf4a	VPTAPLGLYPGKLKSLXXVKXGSTVSAPNDPXXFXRVLVMLDELGXIKLKDXIXXXXXXX	130 140 150 160 170 180

The complete length ORF4a nucleotide sequence <SEQ ID 219> is:

30	1	ATGAAAACCT	TCTTCAAAC	CCTTTCGCGC	GCCGCACTCG	CGCTCATCCT
	51	CGCCGCCTGC	GGCGGTCAA	AAGATAGCGC	GCCCGCCGCA	TCCGCTTCTG
	101	CGCCGC CGCA	CAACGGCGCG	GCGAANAAG	AAATCGCTTT	CGGCACGACC
	151	GTCCGCGCAT	TCCGCGATAT	GGTCAAAGAA	CANATCTCAAC	CCGAGCTGGA
	201	GAAAAAAGGC	TACACCGTCA	AACTGGTCTGA	GTNTACCGAC	TATGTGCGCN
35	251	CGAATCTGGC	ATTGGCTGAG	GGCGAGTTGG	ACATCAACGT	CTTNCAACAC
	301	ANACNCTATC	TTGACGACTN	CAAAAAANAA	CACAATCTGG	ACATCACCNH
	351	AGTCTTNCAA	GTCCGCGACG	CGCCTTTGGG	ACTGTACCCG	GCCGAGCTGA
	401	AATCGCTGGA	NNAAGTCAA	GANGGCAGCA	CCGTATCCCG	GCCCAACGAC
	451	CCGTNNNACT	TCGNCCGCGT	CTTGGTGATG	CTCGACGAAC	TGGGTTNGAT
40	501	CAAACTCAAA	GACNCGCATCA	NNNNGNNGNN	NNNANNCANA	NNNGANANNH
	551	NNNNANNNNH	NNNNNNNNNN	NNNNNCNNCG	NNNNNNNANH	NNNNNNNNNN
	601	NCGNNTNNNN	NNGCNNNNNT	NNANNNTNNN	NNCNCNNNNH	NNNNNTNNNN
	651	NANNANNAGC	GGCATGAAGC	TGACCGAAGC	CCTGTTCCAA	GAACCGAGCT
	701	TTGCCATATG	CAACTGGTCT	GCCGTCAAAA	CCGCCGACAA	AGACGCCCAA
45	751	TGGCTTAAAG	ACGTAACCGA	GGCTATAAC	TCCGACGCGT	TCAAAGCCTA
	801	CGCGCACAAA	CGCTTCGAGG	GCTACAAATC	CCCTGCCGCA	TGGAATGAAG
	851	GCGCAGCCAA	ATAA			

This is predicted to encode a protein having amino acid sequence <SEQ ID 220>:

```

1      MKTFFKTLSA AALALILAAAC GGQKDSAPAA SASAAADNGA AXKEIVFGTT
51     VGDFGDMVKE XIQPELEKKG YTVKLVEXTD YVRXNLALAE GELDINVXQH
101    XXYLDDXKKXK HNLDLTXVKQ VPTAPLGLYP GKKSLXSLXVK GGSTVSAPND
151    PXXFXRVLVLM LDELGXIKLK DXIXXXXXXX XXXXXXXXXX XXXXXXXXXX
201    XXXXAXXXXXX XXXXXXXXXX GMKLTEALFQ EPSFAYVNWS AVKTTADKDSQ
251    WLKDYTEAYN  SDAFKAYAHK RFEGYKSPAA WNEGAAK*

```

A leader peptide is underlined.

Further analysis of these strain A sequences revealed the complete DNA sequence <SEQ ID 221>:

```

1  ATGAAAACCT TCTTCAAAAC CCTTTCGCC GCCGCACTCG CGCTCATCCT
51 CGCCGCTGCG GCGGGTCAAA AAGATAGCGC GCCCGCCGCA TCCGCTTCTG
101 CCGCCGCCGA CAACGGCGCG GCGAAAAAAG AAATCGTCTT CGGCACGACC
151 GTCGGCGACT TCGGCGATAT GGTCAAAGAA CAAATCCAAC CCGAGCTGGA
201 GAAAAAAGGC TACACCGTCA AACTGGTCGA GTTACCGAC TATGTGCGCC
251 CGAATCTGGC ATTGGCTGAG GCGGAGTTGG ACATCAACGT CTTCCAACAC
301 AAACCTATC TTGACGACTT CAAAAAAGAA CACAATCTGG ACATCACCAG
351 AGTCTTCCAA GTGCCGACCG CGCCTTTGGG ACTGTACCCG GGCAAGCTGA
401 AATCGCTGGA AGAAGTCAAA GACGGCAGCA CCGTATCCGC GCCCAACGAC
451 CCGTCCAAC TCGCCGCGCT CTTGGTGATG CTCGACGAAC TGGGTTGGAT
501 CAAACTCAA GACGGCATCA ATCCGCTGAC CGCATCCAAA CGCGACATTG
551 CCGAAAACCT GAAAAACATC AAAATCGTCG AGCTTGAAGC CGCGCAACTG
601 CCGCGTAGCC GCGCGGACGT GGATTTTGCC GTCGTCAACG GCAACTACGC
651 CATAAGCAGC GGCATGAAGC TGACCGAAGC CCTGTTCCAA GAACCGAGCT
701 TTGCCTATGT CAAGTGCTCT GCCGTCAAAA CGCCCGACAA AGACAGCCAA
751 TGGCTTAAAG ACGTAACCGA GGCCTATAAC TCCGACGCGT TCAAAGCCTA
801 CGCGCACAAA CGTTCGAGG GCTACAAATC CCCTGCCGCA TGGAATGAAG
851 GCGCAGCCAA ATAA

```

20 This encodes a protein having amino acid sequence <SEQ ID 222; ORF4a-1>:

```

1  MKTFFKTLSA AALALILAAC GGQKDSAPAA SASAAADNGA AKKEIVFGTT
51 VGDFGDMVKE QIQPELEKKG YTVKLVEFTD YVRPNLALAE GELDINVFQH
101 KPYLDDFKKE HNLDTVEVFQ VPTAPLGLYP GKLSLEEVK DGSTVSAPND
151 PSNFARVLVM LDELGWIKLK DGINPLTASK ADIAENLKNI KIVELEAAQL
201 PRSRADVFA VVNGNYAISS GMKLTEALFQ EPSFAYVNWS AVKTADKDSQ
251 WLKDVTEAYN SDAFKAYAHK RFEGYKSPAA WNEGAAK*

```

ORF4a-1 and ORF4-1 show 99.7% identity in 287 aa overlap:

```

10      20      30      40      50      60
30  orf4a-1  MKTFFKTLSAALALILACGGQKDSAPAASASAAADNGAAKKEIVFGTTVGDFGDMVKE
    orf4-1  MKTFFKTLSAALALILACGGQKDSAPAASASAAADNGAAKKEIVFGTTVGDFGDMVKE
10      20      30      40      50      60

35      70      80      90      100     110     120
    orf4a-1  QIQPELEKKG YTVKLVEFTD YVRPNLALAE GELDINVFQHKPYLDDFKKEHNLDTVEVFQ
    orf4-1  QIQAELEKKG YTVKLVEFTD YVRPNLALAE GELDINVFQHKPYLDDFKKEHNLDTVEVFQ
35      70      80      90      100     110     120

40      130     140     150     160     170     180
    orf4a-1  VPTAPLGLYPGKLSLEEVKDGSTVSAPNDPSNFARVLVMLDELGWIKLKDGINPLTASK
    orf4-1  VPTAPLGLYPGKLSLEEVKDGSTVSAPNDPSNFARVLVMLDELGWIKLKDGINPLTASK
40      130     140     150     160     170     180

45      190     200     210     220     230     240
    orf4a-1  ADIAENLNKIKIVELEAAQLPSRADVDFAVNGNYAISSGMKLTEALFQEPSFAYVNWS
    orf4-1  ADIAENLNKIKIVELEAAQLPSRADVDFAVNGNYAISSGMKLTEALFQEPSFAYVNWS
45      190     200     210     220     230     240

50      250     260     270     280
    orf4a-1  AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEGYKSPAAWNEGAAKX
50      250     260     270     280
    orf4-1  AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEGYKSPAAWNEGAAKX
50      250     260     270     280

```

Homology with an outer membrane protein of *Pasteurella haemolitica* (accession q08869).

ORF4 and this outer membrane protein show 33% aa identity in 91aa overlap:

```

                                10          20
lip2.pasha                      MNEFKLLGVALVSALALTACKDEKAQAP----
                                || | :|| || |:| |:| 
ORF4      VXTPNPDGRTPCPSFLFETATTSGENMKTFKTLSAAL--ALILAACGFKKTARPPHPL
                    110          120          130          140          150

                                30          40          50          60          70          80
lip2.pasha -ATTAKTENKAPLKVGVMGTGPEAQMTEVAVKIAKEKYGLDVELVQFTEYTPQNAAALHSKD
                    : :: | : |: :| ::||: : : || | |:|:|:|::|} } :
ORF4      LPPPTTARRKKEIVFGTTVGDFGDMVKEQIQAELEKKGYTVKLVEFTDYVRPNLALAEGE
                    160          170          180          190          200          210

                                90          100          110          120          130          140
lip2.pasha LDANAFQTVPYLEQEVDKRGYKLAIIGNTLVWPIAAYSKKIKNISCLKDGATVAIPNNAS
                    |
ORF4      L.....

```

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF4 shows 93.6% identity over a 94aa overlap with a predicted ORF (ORF4.ng) from *N. gonorrhoeae*:

[illegible]

The complete length ORF4ng nucleotide sequence <SEQ ID 223> was predicted to encode a protein having amino acid sequence <SEQ ID 224>:

40	1	MKTFFKTLST	ASLALIL <u>LAAC</u>	GGQKDSAPAA	SAAAPSDADNG	AAKKEIVFGT
	51	TVGDFGDMVK	EQIQAELEKK	GYTVKLVEFT	DYVRPNLALA	EGDGLNVFQ
	101	HKPYLDDMFK	EHNLDITEAF	QVPTAPLGLY	PGKLSLEEV	KDGSTVSAPN
	151	DPSNFARALV	MLNELGWIKL	KDGINPLTAS	KADTAENLKN	IKIVELEAAQ
	201	LPRSRAVDVF	AVVNGNYAIS	SGMKLTEALF	QEPSFAYVNW	SAVKTADKDS
45	251	QWLKRDVTEAY	NSDAFKAYAH	KRFEGYKYPA	AWNEGAAK*	

Further analysis revealed the complete length ORF4ng DNA sequence <SEQ ID 225> to be:

	1	atgAAAACCT	TCTTCAAAC	cctttccgcc	gccgcaCTCG	CGCTCATCCT
	51	CGCAGCCTGc	ggCggtcaAA	AAGACAGCGC	GCCCGcagcc	tctgcGCGG
50	101	CCCCTTCTGC	CGATAACGgc	gCgGCGAAAA	AAGAAAtcgt	ctTCGCGACG
	151	ACcgtggcg	acttcggcgA	TatggTCAAA	GAGCAAAATCC	AagcCGAgct
	201	gGAGAAAAAA	GgctACACcg	tcAAattggt	cgaatttacc	gactatgtGC
	251	gCCCCGAATCT	GGCATTGGCG	GAGGGCGAGT	TGGACATCAA	CGCTTTCCAA
	301	CACAAACCTT	ATCTTGACGA	TTTCAAAAAA	GAAACAACCC	TGTACATCAC
55	351	CGAAGCCTTC	CAAGTGCCGA	CCGCGCCTTT	GGGACTGTAT	CCGGGCAAAC
	401	TGAATTCGCT	GGAAGAAGTC	AAAGACGGCA	GCACCGTATC	CGCGCCCAac
	451	gACcgtTCCA	ACTTCGCACG	CGCCTTGGTG	ATGCTGAAGC	AACGTGGTTG
	501	GATCAAACCT	AAAGACGGCA	TCAATCCGCT	GACCGCATCC	AAAGCCGACA
	551	TCGCGGAAAA	CCTGAAAAAC	ATCAAAATCG	TCGAGCTTGA	AGCCGCACAA
60	601	CTGCCGCGCA	GCCGCGCCGA	CGTGGATTTT	GCCCTCGTCA	ACGGCAACTA
	651	CGCCATAAGC	AGCGGCTATGA	AGCTGACCGA	AGCCCTGTTC	CAAGAGCCGA

701 GCTTTGCCTA TGTCAACTGG TCTGCCgtcA AAACCGCCGA CAAAGACAGC  
 751 CAATGGCTTA AAGACGTAAC CGAGGCCTAT AACTCCGACG CGTCAAAGC  
 801 CTACGCGCAC AAACGCTTCG AGGGCTACAA ATACCCTGCC GCATGGAATG  
 851 AAGGCGCAGC CAAATAA

5 This encodes a protein having amino acid sequence <SEQ ID 226; ORF4ng-1>:

1 MKTFFKTLSA AALALILAAC GGQKDSAPAA SAAAPSDNG AAKKEIVFGT  
 51 TVGDFGDMVK EQIQAELEKK GYTVKLVEFT DYVRPNLALA EGELDINVFQ  
 101 HKPYLDDFKK EHNLDITEAF QVPTAPLGLY PGKLKSLEEV KDGSTVSAPN  
 151 DPSNFARALV MLNELGIWIKL KDGINPLTAS KADIAENLKN IKIVELEAAQ  
 201 LPRSRADVDF AVVNGNYAIS SGMKLTEALF QEPSFAYVNW SAVKTADKDS  
 251 QWLKDVTEAY NSDAFKAYAH KRFEQYKYP AAWNEGAAK\*

This shows 97.6% identity in 288 aa overlap with ORF4-1:

		10	20	30	40	50	59
15	orf4-1.pep	MKTFFKTL	SAAALALILA	ACGGQKDS	APAAASA-	AADNGAAK	KEIVFGTTVGDFGDMVK
	orf4ng-1	MKTFFKTL	SAAALALILA	ACGGQKDS	APAAASA-	AADNGAAK	KEIVFGTTVGDFGDMVK
		10	20	30	40	50	60
20	orf4-1.pep	60	70	80	90	100	119
	orf4ng-1	EQIQAELE	KKGYTVKL	VEFTDYVR	PNLALAE	GELDINVF	QHKPYLDDFKKEHNLDITEVF
		70	80	90	100	110	120
25	orf4-1.pep	120	130	140	150	160	179
	orf4ng-1	QVPTAPL	GLYPGKL	KSLEEVK	DGSTVS	APNDPSN	FARVLMDELGIWIKLKDGINPLTAS
		130	140	150	160	170	180
30	orf4-1.pep	180	190	200	210	220	239
	orf4ng-1	KADIAEN	LKNIKIV	ELEAAQL	PRSRADV	DFAVVNG	NYAISSGMKLTEALFQEPSFAYVNW
		190	200	210	220	230	240
35	orf4-1.pep	240	250	260	270	280	
	orf4ng-1	SAVKTAD	KDSQWLK	DVTEAY	NSDAFK	AYAHKR	FEQYKYPAAWNEGAAKX
		250	260	270	280		

In addition, ORF4ng-1 shows significant homology with an outer membrane protein from the database:

45	ID	LIP2 PASHA	STANDARD;	PRT;	276 AA.			
	AC	Q08869;						
	DT	01-NOV-1995	(REL. 32, CREATED)					
	DT	01-NOV-1995	(REL. 32, LAST SEQUENCE UPDATE)					
	DT	01-NOV-1995	(REL. 32, LAST ANNOTATION UPDATE)					
50	DE	28.2 KD OUTER MEMBRANE PROTEIN PRECURSOR. . . .						
	SCORES	Init1: 279	Initn: 416	Opt: 494				
	Smith-Waterman score:	494;	36.0% identity in 275 aa overlap					
55	orf4ng-1.pep	MKTFFKTL	SAAAL--	ALILAAC	GGQKDS	APAAASA	AAPSDNG	AAKKEIVFGTTVGDFGDM
	lip2_pasha	MNFKKLL	GVALV	SALALT	TACKDE	KAQAPATTA	---KTENKAPLK---	VGVMTGPEAQM
		10	20	30	40	50		
60	orf4ng-1.pep	60	70	80	90	100	110	
		VKEIQAE	LEKKGY	TVKLVE	FTDYVR	PNLALAE	GELDINVF	QHKPYLDDFKKEHNLDITE
		:: ::		:::::	::::		::	::: ::

	lip2_pasha	TEVAVKIAKEKYGLDVELVQFTEYTPQNAALHSKDL	DANAFQTVPYLEQEVKDRGYKLAI
		60 70 80 90 100 110	
5	orf4ng-1.pep	120 130 140 150 160 170	AFQVPTAPLGLYPGKLKSL
			EEVKDGVSTVSAPNDPSNFARALV
	lip2_pasha	120 130 140 150 160 170	IGNTLVWP
			IAAYSKIKINISELKD
10	orf4ng-1.pep	180 190 200 210 220 230	ASKADIAENLKN
			IKIVELEAAQLPRSRADVDFAV
	lip2_pasha	180 190 200 210 220 230	ATENDIENPNKIKIVQADT
			SLLRMLDDVELAVINNTYAGQAGLSPDKDGI
15	orf4ng-1.pep	240 250 260 270 280 289	YVNWSAVKTADKDSQWLKDVTEAYNSDAFKAY
			AHKKRFEGYKYPAAWNEGAAKX
	lip2_pasha	240 250 260 270	YVNLVVSREDNKDDPRLQTFVKSFT
			EEVFQEQALKLFNGGVVKGW

Based on this analysis, including the homology with the outer membrane protein of *Pasteurella haemolitica*, and on the presence of a putative prokaryotic membrane lipoprotein lipid attachment site in the gonococcal protein, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF4-1 (30kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figures 8A and 8B show, respectively, the results of affinity purification of the His-fusion and GST-fusion proteins. Purified His-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result), Western blot (Figure 8C), FACS analysis (Figure 8D), and a bactericidal assay (Figure 8E). These experiments confirm that ORF4-1 is a surface-exposed protein, and that it is a useful immunogen.

Figure 8F shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF4-1.

### Example 27

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 227>:

1	CCTCGTCGTC	CTCGGCATGC	TCCAGTTTCA	AGGGGCGATT	TACTCCAAGG
51	CGGTGGAACG	TATGCTCGGC	ACGGTCATCG	GGCTGGGCGC	GGGTTTGGGC
101	GTTTTATGGC	TGAACAGCA	TTATTCCAC	GGCAACCTCC	TCTTCTACCT
151	CACCGTCGGC	ACGGCAAGCG	CACTGGCCGG	CTGGGCGGCG	GTCGGCAAAA
201	ACGGCTACGT	CCCTmTGCTG	GCAGGGCTGA	CGATGTGTAT	GCTCATCGGC
251	GACAACGGCA	GCGAATGGCT	CGACAGCGGA	CTCATGCGCG	CCATGAACGT
301	CCTCATCGGC	GyGGCCATCG	CCATCGCCCG	CGCCAACTG	CTGCCGCTGA
351	AATCCACACT	GATGTGGCGT	TTCATGCTTG	CCGACAACCT	GGCCGACTGC
401	AGCAAAATGA	TTGCCGAAAT	CAGCAACGGC	AGGCGCATGA	CCCGCGAACG
451	CCTCGAGGAG	AACATGGCGA	AAATGCGCCA	AATCAACGCA	CGCATGGTCA
501	AAAGCCGAG	CCATCTCGCC	GCCACATCGG	GCGAAAGCTG	CATCAGCCCC

551 GCCATGATGG AAGCCATGCA GCACGCCAC CGTAAATCG TCAACACCAC  
 601 CGAGCTGCTC CTGACCACCG CCGCCAAGCT GCAATCTCCC AAACCAACG  
 651 GCAGCGAAAT CCGGCTGCTT GACCGCCACT TCACACTGCT CCAAC....  
 701 ..... GC AGACACGCC GCGCATCCG  
 5 CATCGACACC GCCATCAACC CCGAACTGGA AGCCCTCGCC GAACACCTCC  
 801 ACTACCAATG GCAGGGCTTC CTCTGGCTCA GCACCGATAT GCGTCAGGAA  
 851 ATTTCCGCC TCGTCATCCT GCTGCAACGC ACCCGCCGCA AATGGCTGGA  
 901 TGCCACGAA CGCCAACACC TCGCCAAAG CCTGCTTGA

This corresponds to the amino acid sequence <SEQ ID 228; ORF8>:

10 1 .....PRRP RHAPVSRGDL LQGGGTYARH GHRAGRGFGR FMAEPALFPR  
 51 QPPLLPHRRH GKRTGRLGGG RQKRLRPXAG RADDVYAHRR QRQRMARQT  
 101 HARHERPHRR GHRHRRRQTA AAEIHTDVAF HACRQPGRLQ QNDCRNQQRQ  
 151 AHDPRTPRGE HGENAPNQRT HGQKQPSPRR HIGRKLHQR HDGSHAARPP  
 201 XNRQHHRAAP DHRQAAISQ TQRQNPAA PPLHTAPN.....Q  
 15 251 TRPPHPRHRH HQPRTGSPRR TPPLPMAGLP LAQHRYASGN FRPRHPAATH  
 301 PPQMAGCPRT PTPAPKPA\*

Computer analysis of this amino acid sequence gave the following results:

#### Sequence motifs

20 ORF8 is proline-rich and has a distribution of proline residues consistent with a surface localization. Furthermore the presence of an RGD motif may indicate a possible role in bacterial adhesion events.

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF8 shows 86.5% identity over a 312aa overlap with a predicted ORF (ORF8.ng) from *N.gonorrhoeae*:

25 orf8ng 1 MDRDDLRLRRPHAPVPRRDLLQGGGTYARYGHRAGRGFGRFMAEPALFPR 50  
 orf8.pep 1 .....PRRPRHAPVSRGDLQGGGTYARHGHRAGRGFGRFMAEPALFPR 44  
 30 orf8ng 51 QPPLLPHRRHGKRTGRLGGGRQKRLRPYVGADDVHAHRRQRQRMARQRP 100  
 orf8.pep 45 QPPLLPHRRHGKRTGRLGGGRQKRLRPXAGRADDVYAHRRQRQRMARQRT 94  
 35 orf8ng 101 DARDERPHRRHRHCRRTAAAEIHTDVAFHACRQPGRLQNDNRNQQRQ 150  
 orf8.pep 95 HARHERPHRRGHRRRQTAATAAEIHTDVAFHACRQPGRMQNDNRNQQRQ 144  
 40 orf8ng 151 AYDARTFGAEYQNAFNQRTGQKQPSPRRHIGRKPHQLHDGSHAARPP 200  
 orf8.pep 145 AHDPRTPRGEHGENAPNQRTGQKQPSPRRHIGRKLHQRHDGSHAARPP 194  
 45 orf8ng 201 QNRQHHRAAPDHRQAAISQTQRQNPAAAPPLHTAPNRPATNRRPHQRQ 250  
 orf8.pep 195 XNRQHHRAAPDHRQAAISQTQRQNPAAAPPLHTAPN.....Q 244  
 50 orf8ng 251 TRPPHPRHRHQPRTGSPRRTPPLPMAGFPLAQHRYASGNFRPRHPATH 300  
 orf8.pep 245 TRPPHPRHRHQPRTGSPRRTPPLPMAGLPLAQHRYASGNFRPRHPAATH 294  
 orf8ng 301 PPQMAGCPRTPTPAKPA\* 319  
 orf8.pep 295 PPQMAGCPRTPTPAKPA\* 313

The complete length ORF8ng nucleotide sequence <SEQ ID 229> is predicted to encode a protein having amino acid sequence <SEQ ID 230>:

1 MDRDDRLRRP RHAPVPRDL LQGGTYARY GHRAGRGRFGR FMAEPALFPR  
 51 QPPLLPDHRH GKRTGRLGGG RQKRLRPYVG GADDVHAHRR QRQRMARQRP  
 101 DARDERPHRR RHRHCRRQTA AAEIHTDVAF HACRQPGRLQ QNDCRNQQRQ  
 151 AYDARTFGAE YGQNAFNQRT HGQKPQPPRR HIGRKPHQPL HDGSHAARPP  
 5 201 QNRQHHRAAP DHRRQAAISQ TQRQRNPAAR PPLHTAPNRP ATNRRPHQRQ  
 251 TRPPHPRHRH HQPRTGSPRR TPPLPMAGFP LAQHQYASGN FRPRHPPATH  
 301 PPQMAGCPRT PTPAPKPA\*

Based on the sequence motifs in these proteins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 28

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 231>:

1 ..GAAATCAGCC TGCAGTCCGA CNACAGGCCG GTTCCGTGN CGAAGCGGCG  
 51 GGATTCGGAA CGTTTTCTGC TGTTGGACGG CGGCAACAGC CGGCTCAAGT  
 101 GGGCGTGGGT GGAACACGGC ACGTTCGCAA CCGTCCGTAG CGCGCCGTAC  
 151 CGCGATTGTG CGCCTTTGGG CGCGGAGTGG GCGGAAAAGG CGGATGGAAA  
 201 TGTCCGCATC GTCGGTTGCG CTGTGTGCGG AGAATTCAAA AAGGCACAAG  
 251 TGCAGGAACA GCTCGCCCGA AAAATCGAGT GGCTGCCGTC TTCCGCACAG  
 301 GCTTT.GGCA TACGCAACCA CTACCGCCAC CCCGAAGAAC ACGGTTCCGA  
 20 351 CCGCTGGTTC AACGCCTTGG GCAGCCGCCG CTTAGCCGC AACGCCTGCG  
 401 TCGTCGTCAG TTGCGGCACG GCGGTAACGG TTGACGCGCT CACCGATGAC  
 451 GGACATTATC TCGGAGA.GG AACCATCATG CCCGTTTCC ACCTGATGAA  
 501 AGAATCGCTC GCCGTCCGAA CCGCCACCT CAACCGGCAC GCCGGTAAGC  
 551 GTTATCCTTT CCCGACCGG..

25 This corresponds to the amino acid sequence <SEQ ID 232; ORF61>:

1 ..EISLRSDXRP VSVXKRRDSE RFLLLDGGNS RLKWAWVENG TFATVGSAPY  
 51 RDLSPGAEW AEKADGNVRI VGCAVCGEFK KAQVQEQLAR KIEWLPSSAQ  
 101 AXGIRNHYRH PEEHGSDFWF NALGSRFRSR NACVVVSCGT AVTVDALTD  
 151 GHYLGXGTIM PGFHLMKESL AVRTANLNRH AGKRYPFPT..

30 Further work revealed the complete nucleotide sequence <SEQ ID 233>:

1 ATGACGGTTT TGAAGCTTTC GCACTGGCGG GTGTTGGCGG AGCTTGCCGA  
 51 CCGTTTGCCG CAACACGTCT CGCAACTGGC GCGTATGGCG GATATGAAGC  
 101 CGCAGCAGCT CAACGGTTTT TGGCAGCAGA TGCCGGCGCA CATACGCGGG  
 151 CTGTTGCGCC AACACGACGG CTATTGGCGG CTGGTGCGCC CATTGGCGGT  
 201 TTTCGATGCC GAAGGTTTGC GCGAGCTGGG GGAAGGTTCG GGTTTTCAGA  
 251 CGGCATTGAA GCACGAGTGC GCGTCCAGCA ACGACGAGAT ACTGGAATTG  
 301 GCGCGGATTG CGCCGACAAA GGCGCACAAA ACCATATGCG TGACCCACCT  
 351 GCAAAGTAAG GGCAGGGGGC GGCAGGGGCG GAAGTGGTCG CACCGTTTGG  
 401 GCGAGTGTCT GATGTTTCACT TTTGGCTGGG TGTTTGACCG GCCGAGTAT  
 40 451 GAGTTGGGTT CGCTGTCGCC GTTGCGGCA GTGGCGTGTC GCGCGCCTT  
 501 GTCGCGTTTA GGTTTGGATG TGCAGATTAA GTGGCCCAAT GATTTGGTTG  
 551 TCGGACGCGA CAAATTGGGC GGCATTCTGA TTGAAACGGT CAGGACGGGC  
 601 GGCAAAACGG TTGCCGTGGT CCGTATCGGC ATCAATTTTG TCCTGCCCAA  
 651 GGAAGTAGAA AATGCCGCTT CCGTGCAATC GCTGTTTCAG ACGGCATCGC  
 701 GGCGGGGCAA TGCCGATGCC GCCGTGCTGC TGGAACGCT GTTGGTGGAA  
 751 CTGGACGCGG TGTGTTGCA ATATGCGCGG GACGGATTG CGCCTTTTGT  
 801 GCGGGAATAT CAGGCTGCCA ACCGCGACCA CGGCAAGGCG GTATTGCTGT  
 851 TGCGCGACGG CGAAACCGTG TTCGAAGGCA CGGTAAAGG CGTGGACGGA  
 901 CAAGGCGTTT TGCACTTGGA AACGGCAGAG GGCAACAGA CGGTCGTCAG  
 50 951 CCGCGAAATC AGCCTGCGGT CCGACGACAG GCCGTTTCC GTGCCGAAGC  
 1001 GGCGGGATTG GGAACGTTTT CTGCTGTTGG ACGGCGGCAA CAGCCGGCTC  
 1051 AAGTGGGCGT GGGTGGAAAA CGGCACGTTT GCAACCGTCG GTAGCGCGCC  
 1101 GTACCGCGAT TTGTCGCCTT TGGGCGCGGA GTGGGCGGAA AAGGCGGATG  
 1151 GAAATGTCCG CATCGTCGGT TCGCTGTGTG GCGGAGAATT CAAAAGGCA  
 55 1201 CAAGTGCAGG AACAGCTCGC CCGAAAAATC GAGTGGCTGC CGTCTTCCGC  
 1251 ACAGGCTTTG GGCATACGCA ACCACTACCG CCACCCCGAA GAACACGGTT  
 1301 CCGACCGCTG GTTCAACGCC TTGGGCAGCC GCCGCTTCAG CCGCAACGCC

5

1351	TGCGTCGTCG	TCA GTTGC GG	CACGGCGGTA	ACGGTTGACG	CGCTCACCGA
1401	TGACGGACAT	TATCTCGGG	GAACCATCAT	GCCCGGTTTC	CACCTGATGA
1451	AAGAATCGCT	CGCGTCCGA	ACCGCAACC	TC AACCGGCA	CGCCGGTAA G
1501	CGTTATCCTT	TCCCGACCAC	AACGGGCAAT	GCCGTCGCCA	CGCGCATGAT
1551	GGATGCGGTT	TGCGGCTCGG	TTATGATGAT	GCACGGGCGT	TTGAAAGAAA
1601	AAACCGGGGC	GGGCAAGCCT	GTCGATGTCA	TCATTACCGG	CGGCGGCGCG
1651	GCAAAAGTTG	CCGAAGCCCT	GCCGCCGTGA	TTTTTGCGGG	AAAATACCGT
1701	GCGCGTGGCG	GACAACCTCG	TCATTTACGG	GTTGTTGAAC	ATGATTGCCG
1751	CCGAAGGCAG	GG AATATGAA	CATATTTAA		

10 This corresponds to the amino acid sequence <SEQ ID 234; ORF61-1>:

	1	MTVLKLSHRW	VLAELADGLP	QHVSQALARMA	DMKPQQLNGF	WQQMPAHIRG
	51	LLRQHDGYWR	LVRPLAVFDA	EGLRELGERS	GFQTALKHFC	ASSNDEILEL
	101	ARIAPDKAHK	TICVTHLQSK	GRGRQGRKWS	HLRGECLMFS	FGWVFDPRQY
15	151	ELGSLSPVAA	INCFRRALSRL	GLDVQIKWPN	DLVVGDRDKLG	GILIEITRVQT
	201	GKTVAVVGIG	VACRLPKEVE	NAASVQSLFQ	TASRRGNADA	AVLLETLLVE
	251	LDAVLHQYAR	DGFAPFVAEY	QAANRDHGKA	VLLLRDGETV	FEGTVKGVVDG
	301	QGVLLHLETAE	GKQTVVSGEI	SLRSDDRPVS	VPKRRDSERF	LLLDGNGSRL
	351	KWAWVENGTF	ATVGSAPYRD	LSPLGAEWAE	KADGNVRIVG	CAVCGEEFKKA
	401	QVQEQLARKI	EWLPSAQAL	GIRNHRYHPE	EHGSDRWfNA	LGSRRFRSNA
20	451	CVVVSCGTAV	TVDALTDGHH	YLGGTIMPGF	HLMKESLAVR	TANLNHRHAGK
	501	RYPFVPTTTGN	AVASGMMDAV	CGSVMMMHGR	LKEKTGAGRK	VDVITGGGA
	551	AKVAEALPPA	FLAENTVRVA	DNLVIYGLLN	MIAAEGREYE	HI*

Figure 9 shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF61-1. Further computer analysis of this amino acid sequence gave the following results:

25 Homology with the baf protein of *B. pertussis* (accession number U12020).

ORF61 and baf protein show 33% aa identity in 166aa overlap:

```

orf61 23 LLLDGGNSRLKQWAVE-NGTFATVGSAPYR----DLSPLGAEWAEKADGNVRIVGCAVCG 77
      +L+D GNSRLK W + + A AP DL LG A R +G V G
baf 3 ILIDSGNSRLKVGWFDPDAPQAAREPAPVAFDNLDDLALGRWLATLPRRPQRALGVNVAG 62

orf61 78 EFKKAQVQEQLAR---KIEWLPSSAQXGIRNHYRHPEEHGSDRW---FNALGSRRFRSN 131
      + + L I WL + A G+RN YR+P++ G+DRW L +
baf 63 LARGEAIAATLRAGGCDIRWLRQAQLAMGLRNGYRNPDLGADRWACMVGVLARQPSVHP 122

orf61 132 ACVVVSCGTAVTVDALTDGHHYLGXGTIMPGFHLMKESLAVRTANL 177
      +V S GTA T+D + D + G G I+PG +M++ +LA TA+L
baf 123 PLLVASFGTATTLDITIGPDNVFPG-LILIPGAMMRGALAYGTAHL 167

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

40 ORF61 shows 97.4% identity over a 189aa overlap with an ORF (ORF61a) from strain A of *N. meningitidis*:

								10	20	30
	orf6l.pep						EISLRSDXRVPVSVXKRRDSEIFLLLDGGSN			
45	orf6la	TVFEGT	VKGVDGGQVLHLETAEGKQT	VVSGEI	SLSRDDRFPVSPVKRRDSEIFLLLDGGSN					
		290	300	310	320	330	340			
50	orf6l.pep	RLKWAWVENGT	FATVGSA	PYRDLSPLGAEWAEKADGNVRIVGC	AVCGE	FKKAQVQEQLAR				
	orf6la	RLKWAWVENGT	FATVGSA	PYRDLSPLGAEWAEKV	DGNVRIVGC	AVCGE	FKKAQVQEQLAR			
		350	360	370	380	390	400			
55	orf6l.pep	KIEWLPSSAQA	XGIRNH	YRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVT	VDALTDD					
	orf6la	KIEWLPSSAQA	LGI	RNH	YRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVT	VDALTDD				
		410	420	430	440	450	460			



```

              160      170      180      189
orf61.pep    GHYLGXGTIMPGFHLMKESLAVRTANLNRHAGKRYPFPT
              |||||  |||||  |||||  |||||  |||||  |||||
5  orf61a     GHYLG-GTIMPGFHLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDDAVCGSVMMM
              470      480      490      500      510      520

orf61a       HGRLKEKTGAGKPVVDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGG
              530      540      550      560      570      580

```

10 The complete length ORF61a nucleotide sequence <SEQ ID 235> is:

	1	ATGACGCTTT	TGAAGCCTTC	GCACTGGCGG	GTGTTGGCGG	AGCTTGCCGA
	51	CGGTTTGCCG	CAACAGTCT	CGCAACTGGC	GCGTATGGCG	GATATGAAGC
	101	CTGCAGCAGCT	CAACGGTTTT	TGGCAGCAGA	TGCCGGCGCA	CATACCGCGG
15	151	CTGTTTGC CG	AACACGACGG	CTATTGGCGG	CTGGTGC CG	CATTGGCGGT
	201	TTTCGATGCC	GAAGGTTTGC	GCGAGCTGGG	GGAAGGTCG	GGTTTTCAGA
	251	CGGCATTGAA	GCACGAGTGC	GCGTCCAGCA	ACCACGAGAT	ACTGGAATTG
	301	GCGCGGATTG	GCCCGACAA	GGCGCACA	ACCATATGTG	TGACCCACCT
	351	GCAAAGTAA	GGCAGGGGCG	GGCAGGGGCG	GAGTGGTCG	CACCGTTTGG
20	401	GCGAGTGTCT	GATGTTCACT	TTTGGCTGGG	TGTTTGACCG	GCCGCGAGTAT
	451	GAGTTGGGTT	CGCTGTCGCC	TGTTGCGGCA	GTGGCGTGCC	GGCGCGCCTT
	501	GTCGCGT TTT	GTTGTTGAAA	CGCAAATCAA	GTGGCCAAAT	GATTTGGTCG
	551	TCGGACGCGA	CAAATTGGGC	GGCATTCTGA	TTGAAACGGT	CAGGACGGGC
	601	GGCAAAACGG	TTGCCGTGGT	CGGTATCGGC	ATCAATTTTCG	TGCTGCCCAA
25	651	GGAAGTGGAA	AACGCCGCTT	CCGTGCAATC	GCTGTTTCAG	ACGGCATCCG
	701	GCGCGGGGAA	TGCCGATGCC	GCCGTGTTGC	TGGAAACCGT	GTTTGGCGGAA
	751	CTTGATGCGG	TGTTGTTGCA	ATATGCGCGG	GACGGATTTC	CGCCTTTTGT
	801	GGCGGAAATAT	CAGGCTGCCA	ACCGCGACCA	CGGCAAGGCG	GTATTGCTGT
	851	TGCGCGACGG	CGAAACCGTG	TTCGAAGGCA	CGGTAAAGG	CGTGACGCTG
30	901	CAAGGCGTTC	TGCACTTGGA	AACGGCAGAG	GGCAAACAGA	CGGTGCTCAG
	951	CGGCGAAATC	AGCCTGCGGT	CCGACGACAG	GCCGGTTTCC	GTGCCGGAAG
	1001	GGCGGGATTG	GGAACTGTTT	CTGCTGTTGG	ACGGCGGCAA	CAGCCGGCTC
	1051	AAGTGGGCGT	GGGTGGAAAA	CGCACGTTTC	GCAACCGTCG	TGAGCGCGCC
	1101	GTACCGCGAT	TTGTCGCCTT	TGGGCGCGGA	GTGGGCGGAA	AAGGTGGATG
35	1151	GAAATGTCCG	CATCGTCGGT	TGCGCCGTGT	GCGGAGAATT	CAAAAAGGCA
	1201	CAAGTGCAGC	AACAGCTCGC	CCGAAAAATC	GAGTGGCTCG	CTGCTTTCCG
	1251	ACAGGCTTTG	GGCATAACGA	ACCACTACCG	CCACCCCGAA	GAACACGGTT
	1301	CCGACCGCTG	GTTCAACGCC	TTGGGCAGCC	GCCGCTTCAG	CCGCACCGCC
	1351	TGCGTCGCTG	TCACTTGCGG	CACGGCGGTA	ACGGTTGACG	CGCTCACCCA
40	1401	TGACGGACAT	TATCTCGGGG	GAACCATCAT	GCCGGGTTTC	CACCTGATGA
	1451	AAGAAATCGT	CGCCGTCCGA	ACCGCCAACC	TCAACCGGCA	CGCCGGTAAG
	1501	CGTTATCCTT	TCCCGACACC	AACGGGCAAT	GCGCTGCCCA	CGCGCATGAT
	1551	GGATCGGGTT	TGCGGCTCGG	TTATGATGAT	GACGGGCGGT	TTGAAAGAAA
	1601	AAACCGGGGC	GGGCAAGCCT	GTGATGTCTA	TCATTACCGG	CGGCGGCGCG
45	1651	GCAAAAGTTG	CCGAAGCCCT	GCCGCCTGCA	TTTTTGCGCG	AAAATACCGT
	1701	GCGCGTGGCG	GACAACCTCG	TCATTACCGG	GCTGCTGAAC	CTGATTGCCC
	1751	CCGAAGGCGG	GGAAATCGGA	CATACTTAA		

This encodes a protein having amino acid sequence <SEQ ID 236>:

	1	MTVLKPSHR	VLAELADGLP	QHVSQALARMA	DMKPQQLNGF	WQQMPAHIRG
50	51	LLRQHDGYWR	LVRPLAVFDA	EGLRELGERS	GFQTALKHES	ASSNDEILEL
	101	ARIAPDKAHK	TICVTHLQSK	GRGRQGRKWS	HLRVECLMFS	FGWVFDPRQY
	151	ELGSLSPVAA	VACRRALSRL	GLKTQIKWPN	DLVHGCDKLG	GILJETVRTG
	201	GKTVAVVGIG	INFVLPKEVE	NAASVQSLFQ	TASRRGNADA	AVLLETLLAE
	251	LDAVLLQYAR	DGFAPFVAEY	QANRDHGKA	VLLLRDGETV	FEGTVGVGDG
55	301	QGVHLLETAE	GKQTVSGEI	SLRSDDRPVS	VPKRRDSERF	LLLDGGNSRL
	351	KWAVENGTF	ATVGSAPYRD	LSPLGAEWAE	KVDGNVRIVG	CAVCGEFKKA
	401	QVQEQLARKI	EWLPSAQAAL	GIRNHYRHPE	EHGSDRWfNA	LGSRRFRSNA
	451	CVVVSCGTAV	TVDALTTDGH	YLGGTIMPGF	HLMKESLAVR	TANLNRHAGK
	501	RYFPFTTTGN	AVASGMMDAV	CGSVMMHHRG	LKEATGAGKP	VDVITGGGA
	551	AKVAEALPPA	FLAENTVRVA	DNLVIHGLLN	LIAAEGGESE	HT*

60 ORF61a and ORF61-1 show 98.5% identity in 591 aa overlap:

orf61a.pep                10              20              30              40              50              60  
MTVLKPSHWRVLAELADGLPQHVSQQLARMADMKPQQLNQFWQQMPAHIRGLLRQHDGYWR

	orf61-1	MTVLKLSHWRVLAELADGLPQHVSQQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR	10	20	30	40	50	60
			70	80	90	100	110	120
5	orf61a.pep	LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK						
	orf61-1	LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK						
			70	80	90	100	110	120
10	orf61a.pep	GRGRQGRKWSHRLGECMLFSGWVFDROPYELGSLSPVAACRRALSRLGLKTQIKWPN	130	140	150	160	170	180
	orf61-1	GRGRQGRKWSHRLGECMLFSGWVFDROPYELGSLSPVAACRRALSRLGLDVQIKWPN						
			130	140	150	160	170	180
15	orf61a.pep	DLVVGRDKLGGILITVVRTGGKTVAVVGIGINFLPKVEENAASVQSLFQTASRRGNADA	190	200	210	220	230	240
	orf61-1	DLVVGRDKLGGILITVVRTGGKTVAVVGIGINFLPKVEENAASVQSLFQTASRRGNADA						
20			190	200	210	220	230	240
	orf61a.pep	AVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHKGAVLLLRDGETVFEGTVKGVDG	250	260	270	280	290	300
25	orf61-1	AVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHKGAVLLLRDGETVFEGTVKGVDG						
			250	260	270	280	290	300
30	orf61a.pep	QGVHLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSEFLLLDGGNSRLKWAWVENGTF	310	320	330	340	350	360
	orf61-1	QGVHLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSEFLLLDGGNSRLKWAWVENGTF						
			310	320	330	340	350	360
35	orf61a.pep	ATVGSAPYRDLSPGAEWAEEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL	370	380	390	400	410	420
	orf61-1	ATVGSAPYRDLSPGAEWAEEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL						
			370	380	390	400	410	420
40	orf61a.pep	GIRNHYRHPEEHGSDRWFNALGSRFRSACVNVVSCGTAVTVDALTDGHHYLGGTIMPGE	430	440	450	460	470	480
	orf61-1	GIRNHYRHPEEHGSDRWFNALGSRFRSACVNVVSCGTAVTVDALTDGHHYLGGTIMPGE						
45			430	440	450	460	470	480
	orf61a.pep	HLMKESLAVRTANLNRHAGKRYFPPTTTGNAVASGMMDAVCGSVMMHGRLEKTGAGKP	490	500	510	520	530	540
	orf61-1	HLMKESLAVRTANLNRHAGKRYFPPTTTGNAVASGMMDAVCGSVMMHGRLEKTGAGKP						
50			490	500	510	520	530	540
	orf61a.pep	VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAEGGESEHTX	550	560	570	580	590	
55	orf61-1	VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIYGLLNLMIAAEGREYEHIX						
			550	560	570	580	590	

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF61 shows 94.2% identity over a 189aa overlap with a predicted ORF (ORF61.ng) from *N.*

60 *gonorrhoeae*:

orf61.pep	EISLRSDXRPVSVXKRRDSEFLLLDGGNS	30
orf61ng	TVCEGTVKGVDGRGVLHLETAEGEQTVVSGEISLRPDNRSVSVPKRPDSEFLLLEGGNS	211

	orf61.pep	RLKWAWVENGTTFATVGSAPYRDLSPGLAEWAEEKADGNVRIVGCAVCGEFKKAQVQEQLAR	90
	orf61ng	RLKWAWVENGTTFATVGSAPYRDLSPGLAEWAEEKADGNVRIVGCAVCGESKKAQVKEQLAR	271
5	orf61.pep	KIEWLPSSAQAXGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTD	150
	orf61ng	KIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTD	331
10	orf61.pep	GHYLGXGTIMPGFHLMKESLAVRTANLNRHAGKRYFPPT	189
	orf61ng	GHYLG-GTIMPGFHLMKESLAVRTANLNRPAKRYFPPTTTGNAVASGMMDAVCGSIMMM	390

An ORF61ng nucleotide sequence <SEQ ID 237> was predicted to encode a protein having amino acid sequence <SEQ ID 238>:

15	1	MFSEGWAFDR	PQYELGSLSP	VAALACRRAL	GCLGLETQIK	WPNDLVVGRD
	51	KLGGILIETV	RAGGKTAVV	GIGINFVLPK	EVENAASVQS	LFQTASRRGN
	101	ADAAVLLETL	LAELGAVLEQ	YAEEGFAPFL	NEYETANRDH	GKAVLLLRDG
	151	ETVCEGTVK	VDGRGVHLE	TAEGEQTVVS	GEISLRPDNR	SVSVKRPDS
	201	ERFLLEGGN	SRLKWAVVEN	GTATVGSAP	YRDLSPGLAE	WAEKADGNVR
20	251	IVGCAVCGES	KKAQVKEQLA	RKIEWLPSSA	QALGIRNHYR	HPEEHGSDRW
	301	FNALGSRRFS	RNACVVVSCG	TAVTVDALTD	DGHYLGGTIM	PGFHLMKESL
	351	AVRTANLNR	AGKRYFPPTT	TGNAVASGMM	DAVCGSIMMM	HGRLKEKNGA
	401	GKPVVDVITG	GGAAKVAEAL	PPAFLAENTV	RVADNLVIHG	LLNLIAAEGG
	451	ESEHA*				

Further analysis revealed the complete gonococcal DNA sequence <SEQ ID 239> to be:

25	1	ATGACGGTTT	TGAAGCCTTC	GCATTGGCGG	GTGTTGGCGG	AGCTTGCCGA
	51	CGGTTTGCCG	CAACACGTAT	CGCAATTGGC	GCGTGAGGCG	GACATGAAGC
	101	CGCAGCAGCT	CAACGGTTTT	TGGCAGCAGA	TGCCGCGCGA	TATACGCGGG
	151	CTGTGCGGCC	AACACGACGG	CTATTGGCGG	CTGGTGGCGC	CCTTGGCGGT
	201	TTTCGATGCC	GAAGGTTTGC	GCGATCTGGG	GGAAAGGTCG	GGTTTTCAGA
30	251	CGGCATTGAA	GCACGAGTGC	GCGTCCAGCA	ACGACGAGAT	ACTGGAATTG
	301	GCGCGGATTG	CGCCGGACAA	GGCGCACAAA	ACCATATGCG	TGACCCACCT
	351	GCAAAGTAAG	GGCAGGGGGC	GGCAGGGGGC	GAAGTGGTCG	CACCGTTTGG
	401	GCGAGTGCCT	GATGTTTCAGT	TTCGGCTGGG	CGTTTGACCG	GCCGCGATAT
	451	GAGTTGGGTT	CGCTGTGCGC	TGTTGCGGCA	CTTGCGTGCC	GGCGCGCTTT
35	501	GGGGTGTTTG	GGTTTGAAA	CGCAAATCAA	GTGGCCAAAC	GATTTGGTTCG
	551	TCCGACGCGA	CAAAATGGGC	GGCATTCTGA	TTGAAACAGT	CAGGGCGGGC
	601	GGTAAACCGG	TTGCCGTGGT	CGGTATCGGC	ATCAATTTCG	TGCTGCCCAA
	651	GGAAAGTGAA	AACGCGGCTT	CCGTGCAGTC	GCTGTTTCAG	ACGGCATCCG
40	701	GGCGGGGCAA	TGCCGATGCC	GCCGTATTGC	TGGAACATT	GCTTGC GGAA
	751	CTGGGCGCGG	TGTTGGAACA	ATATGCGGAA	GAAGGGTTCG	CGCCATTTTT
	801	AAATGAGTAT	GAAACGGCCA	ACCGCGACCA	CGGCAAGGCG	GTATTGCTGT
	851	TGCGCGACGG	CGAAACCGTG	TGCGAAGGCA	CGGTAAAGG	CGTGGACGGA
	901	CGAGGCGTTC	TGCATTGGA	AACGGCAgaa	ggcgaACAGa	cggtcgtcag
	951	cggcgaaaTC	AGcctGCggc	ccgacaacaG	GTCGGtttcc	gtgccgaagc
45	1001	ggcgggatTC	GgaacgtTTT	tTGCTgttgg	aagcggggaa	cagccgGCTC
	1051	AAGTGGGCGT	GggtggAAAa	cggcacgttc	gcaaccgtgg	gcagcgcgCc
	1101	gtaCCGCGAT	TTGTGCGCTT	TGGGCGCGGA	GTGGGCGGAA	AAGGCGGATG
	1151	GAAATGTCCG	CATCGTCGGT	TGCGCCGTGT	GCGGAGAATC	CAAAAAGGCA
50	1201	CAAGTGAAGG	AACAGCTCGC	CCGAAAAATC	GAGTGGCTGC	CGTCTTCCGC
	1251	ACAGGCTTTG	GGCATACGCA	ACCACTACCG	CCACCCGAA	GAACACGGTT
	1301	CCGACCGTTG	GTTCAACGCC	TTGGGCAGCC	GCCGCTTCAG	CCGCAACGCC
	1351	TGCGTCGTCG	TCAGTTGCGG	CACGGCGGTA	ACGGTTGACG	CGCTCACCGA
	1401	TGACGGACAT	TATCTCGGCG	GAACCATCAT	GCCCGGCTTC	CACCTGATGA
55	1451	AAGAATCGCT	CGCCGTCGGA	ACCGCCAACC	TCAACCGCCC	CGCCGGGAAA
	1501	CGTTACCCTT	TCCCGACCAC	AACGGGCAAC	GCCGTCGCAA	GCGGCATGAT
	1551	GGACGCGGTT	TCCGGCTCGA	TAATGATGAT	GCACGGCCGT	TTGAAAGAAA
	1601	AAAACGGCGC	GGGCAAGCCT	GTCGATGTCA	TCATTACCGG	CGGCGGCGCG
	1651	GCGAAAGTCG	CCGAAGCCCT	GCCGCCTGCA	TTTTTGGCGG	AAAATACCGT
	1701	GCGCGTGGCG	GACAACCTCG	TCATCCACGG	GCTGCTGAAC	CTGATTGCCG
60	1751	CCGAAGGCGG	GGAATCGGAA	CACGCTTAA		

This corresponds to the amino acid sequence <SEQ ID 240; ORF61ng-1>:

1 MTVLKPSHWR VLAELADGLP QHVSQ LAREA DMKPQQLNGF WQMPAHIRG

51 LLRQHDGYWR LVRPLAVFDA EGLRDLGERS GFQTALKHEC ASSNDEILEL  
 101 ARIAPDKAHK TICVTHLQSK GRGRQGRKWS HRLGECLMFS FGWAFDRPQY  
 151 ELGSLSPVAA LACRRALGCL GLETQIKWPN DLVVGRDKLG GILIETVRAG  
 201 GKTVAUVGIG INFVLPKEVE NAASVQSLFQ TASRRGNADA AVLLETLLAE  
 5 LGAVLEQYAE EGFAPFLNEY ETANRDHGKA VLLLRDGETV CEGTVKGVDG  
 301 RGV LHLETAE GEQTVVSGEI SLRPDNRSVS VPKRPDSERF LLEGGNSRL  
 351 KWAUVENGTF ATVGSAPYRD LSPLGAEWAE KADGNVRIVG CAVCGESKKA  
 401 QVKEQLARKI EWLPSSAQAL GIRNHYRHPE EHGS DRWFNA LGSRRFSRNA  
 451 CVVVSCGTAV TVDALTDGHH YLGGTIMPGF HLMKESLAVR TANLNRPAKG  
 10 RYFPPTTGN AVASGMMDAV CGSIMMHGR LKEKNGAGKP VDVIIITGGGA  
 551 AKVAEALPPA FLAENTVRVA DNLVIHGLLN LIAAEGGESE HA\*

ORF61ng-1 and ORF61-1 show 93.9% identity in 591 aa overlap:

15 orf61ng-1.pep MTVLKPSHWRVLAELADGLPQHVSQIAREADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60  
 orf61-1 MTVLKLSHWRVLAELADGLPQHVSQIARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60  
 20 orf61ng-1.pep LVRPLAVFDAEGLRDLGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120  
 orf61-1 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120  
 25 orf61ng-1.pep GRGRQGRKWSHRLGECLMFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPN 180  
 orf61-1 GRGRQGRKWSHRLGECLMFSFGWVDFRPQYELGSLSPVAACRRALSRLGLDVQIKWPN 180  
 30 orf61ng-1.pep DLVVGRDKLGGILIETVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240  
 orf61-1 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240  
 35 orf61ng-1.pep AVLLETLLAE LGAVLEQYAE EGFAPFLNEY ETANRDHGKA VLLLRDGETV CEGTVKGVDG 300  
 orf61-1 AVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKA VLLLRDGETV FEETVKGVDG 300  
 40 orf61ng-1.pep RGV LHLETAEGEQTVVSGEISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAUVENGTF 360  
 orf61-1 QGV LHLETAEGEQTVVSGEISLRSDRPVSVPKRRDSERFLLLDGGNSRLKWAUVENGTF 360  
 45 orf61ng-1.pep ATVGSAPYRDLSPPLGAEWAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQAL 420  
 orf61-1 ATVGSAPYRDLSPPLGAEWAEKADGNVRIVGCAVCGEFKKAQVQEQQLARKIEWLPSSAQAL 420  
 50 orf61ng-1.pep GIRNHYRHPEEHGSDRWFNALGSRFRSRNACVVVSCGTAVTVDALTDGHHYLGGTIMPGF 480  
 orf61-1 GIRNHYRHPEEHGSDRWFNALGSRFRSRNACVVVSCGTAVTVDALTDGHHYLGGTIMPGF 480  
 55 orf61ng-1.pep HLMKESLAVRTANLNRPAKGKRYFPPTTGN AVASGMMDAVCGSIMMHGR LKEKNGAGKP 540  
 orf61-1 HLMKESLAVRTANLNRHAGKRYFPPTTGN AVASGMMDAVCGSVMMMHGR LKEKTGAGKP 540  
 orf61ng-1.pep VDVIIITGGGA AKVAEALPPA FLAENTVRVADNLVIHGLLN LIAAEGGESEHAX 593  
 orf61-1 VDVIIITGGGA AKVAEALPPA FLAENTVRVADNLVIYGLLN MIAAEGREYEHIX 593

Based on this analysis, including the homology with the baf protein of *B. pertussis* and the presence of a putative prokaryotic membrane lipoprotein lipid attachment site, it is predicted that these proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 29

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 241>:

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60

```

1  ATGTTTACC AAATCCTTGC CCTGATTATC TGGAGCAGCT CGTTTATTGC
51 CGCCAAATAT GTCTATGGCG GCATCGATCC CGCATTGATG GTCGGCGTGC
101 GCCTGCTAAT TGCCGCGCTG CCTGCACTGC CCGCCTGCCG CCGTCATGTC
151 GGCAAGATTC CGCGTGAGGA ATGGAAGCCG TTGCTGATTG TGTCGTTTCGT
201 CAACTATGTG CTGACCCTGC TGCTTCAGTT TGTCGGGTTG AAATACACTT
251 CCGCCGCCAG CGCATCGGTC ATTGTCGGAC TCGAGCCGCT GCTGATGGTG
301 TTTGTCGGAC ACTTTTCTT CAACGACAAA GCGCGTGCCT ACCACTGGAT
351 ATGCGGCGCG GCGGCATTG CCGGTGTCGC GCTGCTGATG GCGGGCGGTG
401 CGGAAGAGGG GCGCGAAGTC GGCTGGTTCG GCTGCCTGCT GGTGTTGTG
451 GCGGGCGCGG GCTTTTGTG CGCTATGCGT CCGACGCAA GGCTGATTGC
501 ACGCATCGGC GCACCGGCAT TCACATCTGT TTCCATTGCC GCCGCATCGT
551 TGATGTGCCT GCCGTTTTCG CTGCTTTGG CGCAAAGTTA TACCGTGGAC
601 TGGAGCGTCG GGATGGTATT GTCGCTGCTG TATTTGGGTT TGGGGTGC..

```

This corresponds to the amino acid sequence <SEQ ID 242; ORF62>:

15  
20  
25  
30  
35  
40  
45  
50  
55  
60

```

1  MFYQILALII WSSSFIAAKY VYGGIDPALM VGVRLIIAAL PALPACRRHV
51 GKIPREEWKP LLIVSFVNYV LTLLLQFVGL KYTSAASASV IVGLEPLLMV
101 FVGHHFFNDK ARAYHWICGA AAFAGVALLM AGGAEEGGEV GWFGCLLVLL
151 AGAGFCAAMR PTQRLIARIG APAFTSVSIA AASLMCLPFS LALAQSYTVD
201 WSVGMVLSLL YLGLGCGWYA YLWLNKMSR VPANVSGLLI SLEPVVGVLL

```

20 Further work revealed the complete nucleotide sequence <SEQ ID 243>:

25  
30  
35  
40  
45  
50  
55  
60  
65  
70  
75  
80  
85

```

1  ATGTTTACC AAATCCTTGC CCTGATTATC TGGAGCAGCT CGTTTATTGC
51 CGCCAAATAT GTCTATGGCG GCATCGATCC CGCATTGATG GTCGGCGTGC
101 GCCTGCTAAT TGCCGCGCTG CCTGCACTGC CCGCCTGCCG CCGTCATGTC
151 GGCAAGATTC CGCGTGAGGA ATGGAAGCCG TTGCTGATTG TGTCGTTTCGT
201 CAACTATGTG CTGACCCTGC TGCTTCAGTT TGTCGGGTTG AAATACACTT
251 CCGCCGCCAG CGCATCGGTC ATTGTCGGAC TCGAGCCGCT GCTGATGGTG
301 TTTGTCGGAC ACTTTTCTT CAACGACAAA GCGCGTGCCT ACCACTGGAT
351 ATGCGGCGCG GCGGCATTG CCGGTGTCGC GCTGCTGATG GCGGGCGGTG
401 CGGAAGAGGG GCGCGAAGTC GGCTGGTTCG GCTGCCTGCT GGTGTTGTG
451 GCGGGCGCGG GCTTTTGTG CGCTATGCGT CCGACGCAA GGCTGATTGC
501 ACGCATCGGC GCACCGGCAT TCACATCTGT TTCCATTGCC GCCGCATCGT
551 TGATGTGCCT GCCGTTTTCG CTGCTTTGG CGCAAAGTTA TACCGTGGAC
601 TGGAGCGTCG GGATGGTATT GTCGCTGCTG TATTTGGGTT TGGGGTGC
651 CTGGTACGCC TATTGGCTGT GGAACAAGGG GATGAGCCGT GTTCCTGCCA
701 ATGTTTCGGG ACTGTTGATT TCGCTCGAAC CCGTCGTCGG CGTCTGCTG
751 GCGGTTTTGA TTTTGGGCGA ACACCTGTG CCGTGTCCG CCTTGGGCGT
801 GTTTGTCGTC ATCGCCGCCA CCTTGGTTGC CGGCCGGCTG TCGCATCAA
851 AATAA

```

This corresponds to the amino acid sequence <SEQ ID 244; ORF62-1>:

40  
45  
50  
55  
60  
65  
70  
75  
80  
85

```

1  MFYQILALII WSSSFIAAKY VYGGIDPALM VGVRLIIAAL PALPACRRHV
51 GKIPREEWKP LLIVSFVNYV LTLLLQFVGL KYTSAASASV IVGLEPLLMV
101 FVGHHFFNDK ARAYHWICGA AAFAGVALLM AGGAEEGGEV GWFGCLLVLL
151 AGAGFCAAMR PTQRLIARIG APAFTSVSIA AASLMCLPFS LALAQSYTVD
201 WSVGMVLSLL YLGLGCGWYA YLWLNKMSR VPANVSGLLI SLEPVVGVLL
251 AVLILGEHLS PVSALGVFVV IAATLVAGRL SHQK*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical transmembrane protein HI0976 of *H. influenzae* (accession number Q57147)

ORF62 and HI0976 show 50% aa identity in 114aa overlap:

50  
55

```

Orf62 1 MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLIIAALPALPACRRHV 60
      M YQILAL+IWSSS I K Y +DP L+V VR R KI + K
HI0976 1 MLYQILALLIWSSSLIVGKLTYSMMDPVLVVQVRLIIAMIIVMPLFLRRWKIDKPMRKQ 60

Orf62 61 LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLMVFGHHFFNDKARAY 114
      L ++F NY LLQF+GLKYTSA+SA ++GLEPLL+VFVGHFFF K +
HI0976 61 LWWLAFFNYTAVFLLQFVGLKYTSAASAVTMIGLEPLLVVFGHHFFKTKQNGF 114

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF62 shows 99.5% identity over a 216aa overlap with an ORF (ORF62a) from strain A of *N. meningitidis*:

5	orf62.pep	MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHV	10	20	30	40	50	60
	orf62a	MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHV	10	20	30	40	50	60
10	orf62.pep	LLIVSFVNYVLTLLQLFVGLKYTSAASASVIVGLEP	70	80	90	100	110	120
	orf62a	LLIVSFVNYVLTLLQLFVGLKYTSAASASVIVGLEP	70	80	90	100	110	120
15	orf62.pep	AAFAGVALLMAGGAEEGGEVGF	130	140	150	160	170	180
	orf62a	AAFAGVALLMAGGAEEGGEVGF	130	140	150	160	170	180
20	orf62.pep	AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGC	190	200	210			
	orf62a	AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGVGCSWYAYWLWNKGMSRVPANVSGLLI	190	200	210	220	230	240
25	orf62a	SLEPVVGVLAVLILGEHLS	250	260	270	280		

30 The complete length ORF62a nucleotide sequence <SEQ ID 245> is:

1	ATGTTTACC	AAATCCTTGC	CCTGATTATC	TGGAGCAGCT	CGTTTATTGC
51	CGCCAAATAT	GTCTATGGCG	GCATCGATCC	CGCATGATG	GTCGGCGTGC
101	GCCTGCTGAT	TGCTGCGCTG	CCTGCACTGC	CCGCTGCCC	CCGTCATGTC
151	GGCAAGATTC	CGCGTGAGGA	ATGGAAGCCG	TTGCTGATTG	TGTCGTTCGT
201	CAACTATGTG	CTGACCCTGC	TACTTCAGTT	TGTCGGGTTG	AAATACACTT
251	CCGCCGCCAG	CGCATCGGTC	ATTGTCGGAC	TCGAGCCACT	GCTGATGGTG
301	TTTGTGCGAC	ACTTTTTCTT	CAACGACAAA	GCGCGTGCCT	ACCACTGGAT
351	ATGCGGCGCG	GCGGCATTTC	CCGGTGTCGC	GCTGCTGATG	GCGGGCGGTG
401	CGGAAGAGGG	CGGCGAAGTC	GGCTGGTTCG	GCTGCTGCTG	GGTGTGTTG
451	GCGGGCGCGG	GCTTTTGTGC	CGCTATGCGT	CCGACGAAA	GGCTGATTGC
501	ACGCATCGGC	GCACCGGCAT	TCACATCTGT	TTCCATTGCC	GCCGCATCGT
551	TGATGTGCCT	GCCGTTTTTC	CTTGCTTTGG	CGCAAAGTTA	TACCGTGGAC
601	TGGAGCGTCG	GAATGGTATT	GTCGCTGCTG	TATTGGGCG	TGGGGTGCAG
651	CTGGTACGCC	TATTGGCTGT	GGAACAAGGG	GATGAGCCGT	GTTCTGCCA
701	ACGTTTCGGG	ACTGTTGATT	TCGCTCGAAC	CCGTCGTCG	CGTGCTGCTG
751	GCGGTTTGA	TTTTGGGCGA	ACACCTGTGC	CCCGTGCCG	TCTTGGGCGT
801	GTTTGTGCTC	ATCGCCGCCA	CCTTGGTTGC	CGGCCGGCTG	TCGCATCAA
851	AATAA				

This encodes a protein having amino acid sequence <SEQ ID 246>:

50	1	MFYQILALII	WSSSFIAAKY	VYGGIDPALM	VGVRLLIAAL	PALPACRRHV
	51	GKIPREEWKP	LLIVSFVNYV	LTLLQLFVGL	KYTSASASV	IVGLEP
	101	FVGHFFNDK	ARAYHWICGA	AAFAGVALLM	AGGAEEGGEV	GWFGCLLVLL
	151	AGAGFCAAMR	PTQRRIARIG	APAFTSVSIA	AASLMCLPFS	LALAQSYTVD
	201	WSVGMVLSLL	YLVGVCSWYA	YWLWNKGMSR	VPANVSGLLI	SLEPVVGVL
55	251	AVLILGEHLS	PVSVLGVFVV	IAATLVAGRL	SHQK*	

ORF62a and ORF62-1 show 98.9% identity in 284 aa overlap:

60	orf62a.pep	MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHV	60
	orf62-1	MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHV	60

	orf62a.pep	LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLVMFVGHHFFNDKARAYHWICGA	120
	orf62-1	LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLVMFVGHHFFNDKARAYHWICGA	120
5	orf62a.pep	AAFAGVALLMAGGAEEGGEVGFGLLVLLAGAGFCAAMRPTQRLIARIGAPAFSTVSIA	180
	orf62-1	AAFAGVALLMAGGAEEGGEVGFGLLVLLAGAGFCAAMRPTQRLIARIGAPAFSTVSIA	180
10	orf62a.pep	AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGVGCSWYAYWLWNKGMSRVPANVSGLLI	240
	orf62-1	AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGCGWYAYWLWNKGMSRVPANVSGLLI	240
	orf62a.pep	SLEPVVGVLAVLILGEHLSPVSVLGVFVIAATLVAGRLSHQKX	285
15	orf62-1	SLEPVVGVLAVLILGEHLSPVSVLGVFVIAATLVAGRLSHQKX	285

Homology with a predicted ORF from *N.gonorrhoeae*

ORF62 shows 99.5% identity over a 216aa overlap with a predicted ORF (ORF62.ng) from *N. gonorrhoeae*:

20	orf62.pep	MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVKGIPREEWKP	60
	orf62ng	MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVKGIPREEWKP	60
25	orf62.pep	LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLVMFVGHHFFNDKARAYHWICGA	120
	orf62ng	LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLVMFVGHHFFNDKARAYHWICGA	120
	orf62.pep	AAFAGVALLMAGGAEEGGEVGFGLLVLLAGAGFCAAMRPTQRLIARIGAPAFSTVSIA	180
30	orf62ng	AAFAGVALLMAGGAEEGGEVGFGLLVLLAGAGFCAAMRPTQRLIARIGAPAFSTVSIA	180
	orf62.pep	AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGC	216
	orf62ng	AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGCGWYAYWLWNKGMSRVPANASGLLI	240

35 The complete length ORF62ng nucleotide sequence <SEQ ID 247> is:

	1	ATGTTTTACC	AAATCCTTGC	CCTGATTATC	TGGGGCAGCT	CGTTTATTGC
	51	CGCCAAATAT	GTCTATGGCG	GCATCGATCC	CGCATTGATG	GTCGGCGTGC
	101	GCCTGCTGAT	TGCCCGCGCTG	CCTGCACTGC	CCGCTGCGG	CCGTGATGTC
	151	GGCAAGATTC	CGCGTGAGGA	ATGGAAGCCG	TTGCTGATTG	TGTCGTTCTG
40	201	CAACTATGTG	CTGACCCCTGC	TGCTTCAGTT	TGTCGGGTTG	AAATACACTT
	251	CCGCCGCCAG	CGCATCGGTC	ATTGTCGGAC	TCGAGCCGCT	GCTGATGGTG
	301	TTTGTCGGAC	ACTTTTCTT	CAACGACAAA	GCGCGTGCCT	ACCACTGGAT
	351	ATGCGGCGCG	GCGGCATTTG	CCGCTGTGCG	GCTGCTGATG	GCGGGCGGTG
45	401	CGGAAGAGGG	CGGCGAAGTC	GGCTGGTTCG	GCTGCCTGCT	GCTGTTGTTG
	451	GCGGGCGCGG	GCTTTGTGTC	CGCTATGCGT	CCGACGCAAA	GGCTGATTGC
	501	CCGCATCGGC	GCACCGGCAT	TCACATCTGT	TTCCATTGCC	GCCGCATCGT
	551	TGATGTGCCT	GCCGTTTTCG	CTTGCTTTGG	CGCAAAGTTA	TACCGTGGAC
	601	TGGAGCGTCG	GGATGGTATT	GTCGCTGTTG	TATTTGGGTT	TGGGGTGCGG
	651	CTGGTACGCC	TATTGGCTGT	GGAACAAGGG	GATGAGCCGT	GTTCTCTCCA
50	701	ACGCGTCGGG	ACTGTTGATT	TCGCTCGAAC	CCGTCGTCGG	CGTGCTGTTG
	751	GCGGTTTTGA	TTTTGGGCGA	ACATTTATCG	CCCGTGTCGG	CCTTGGGCGT
	801	GTTTGTGTCG	ATCGCCGCCA	CTTTCGCCGC	CGGCCGGCTG	TCGCGCAGGG
	851	ACGCGCAAAA	CGGCAATGCC	GTCTGA		

This encodes a protein having amino acid sequence <SEQ ID 248>:

55	1	MFYQILALII	WGSSSFIAAKY	VYGGIDPALM	VGVRLLIAAL	PALPACRRHV
	51	GKIPREEWKP	LLIVSFVNYV	LTLQLQFVGL	KYTSASASV	IVGLEPLLVM
	101	FVGHHFFNDK	ARAYHWICGA	AAFAGVALLM	AGGAEEGGEV	GWFGCLLVLL
	151	AGAGFCAAMR	PTQRLIARIG	APAFSTVSIA	AASLMCLPFS	LALAQSYTVD
	201	WSVGMVLSLL	YLGIGCGWYA	YWLWNKGMSR	VPANASGLLI	SLEPVVGVL
60	251	AVLILGEHLS	PVSALGVFV	IAATFAAGRL	SRRDAQNGNA	V*

ORF62ng and ORF62-1 show 97.9% identity in 283 aa overlap:

		10	20	30	40	50	60
	orf62ng.pep	MFYQILALIIWGSSFIAAKYVYGGIDPALMVGVRLIIAALPALPACRRHVKGIPREEWKP					
5	orf62-1	MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLIIAALPALPACRRHVKGIPREEWKP					
		10	20	30	40	50	60
	orf62ng.pep	LLIVSFVNYVLTLLQLFVGLKYTSAASASVIVGLEPLLVMFVGHFFFNDKARAYHWICGA					
10	orf62-1	LLIVSFVNYVLTLLQLFVGLKYTSAASASVIVGLEPLLVMFVGHFFFNDKARAYHWICGA					
		70	80	90	100	110	120
	orf62ng.pep	AAAFAGVALLMAGGAEEGGEVGFVGLVLLAGAGFCAAMRPTQRLIARIGAPFTSVSIA					
15	orf62-1	AAAFAGVALLMAGGAEEGGEVGFVGLVLLAGAGFCAAMRPTQRLIARIGAPFTSVSIA					
		130	140	150	160	170	180
	orf62ng.pep	AASLMCLPFSLALAQSYTVDWVSGMVLSSLYLGLGCGWYAYWLWNKGMSRVPANASGLLI					
20	orf62-1	AASLMCLPFSLALAQSYTVDWVSGMVLSSLYLGLGCGWYAYWLWNKGMSRVPANVSGLLI					
		190	200	210	220	230	240
	orf62ng.pep	SLEPVVGVLAVLILGEHLSPVSALGVFVIAATFAAGRLSRRDAQNGNAVX					
25	orf62-1	SLEPVVGVLAVLILGEHLSPVSALGVFVIAATLVAGRLSHQKX					
		250	260	270	280	290	

Furthermore, ORF62ng shows significant homology to a hypothetical *H. influenzae* protein:

sp|Q57147|Y976\_HAEIN HYPOTHETICAL PROTEIN HI0976 >gi|1074589|pir||B64163  
 hypothetical protein HI0976 - Haemophilus influenzae (strain Rd KW20)  
 >gi|1574004 (U32778) hypothetical [Haemophilus influenzae] Length = 128  
 Score = 106 bits (262), Expect = 2e-22  
 Identities = 56/114 (49%), Positives = 68/114 (59%)

Query: 1 MFYQILALIIWGSSFIAAKYVYGGIDPALMVGVRRXXXXXXXXXXCRRHVKGIPREEWKP 60  
 M YQILAL+IW SS I K Y +DP L+V VR R KI + K

Sbjct: 1 MLYQILALLIWSSSLIVGKLTYSMMDPVLVVQVRLIIAMIIVMPLFLRRWKKIDKPMRKQ 60

Query: 61 LLIVSFVNYVLTLLQLFVGLKYTSAASASVIVGLEPLLVMFVGHFFFNDKARAY 114  
 L ++F NY LLQF+GLKYTSA+SA ++GLEPLL+VEVGHFFF K +

Sbjct: 61 LWWLAFFNYTAVFLLQFVGLKYTSAASAVTMIGLEPLLVMFVGHFFFKTKQNGF 114

Based on this analysis, including the homology with the transmembrane protein of *H. influenzae* and the putative leader sequence and several transmembrane domains in the gonococcal protein, it is predicted that these proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 30

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 249>:

1 ATGCGCCGTT TTCTACCGAT CGCAGCCATA TGCGCmGwms TCCTGkkGTA  
 51 sGGACTGACG GCGGCAACCG GCAGCACCAG TTCGCTGGCG GATTATTTCT  
 101 GGTGGATTGT TCGGTTACAG GCAATGCTGC TGCTGGTGTT GTCCGCCGTT  
 151 TTGGCACGTT ATGTCATATT GCTGTTGAAA GACAGGCGCG ACGGCGTATT  
 201 CGGTTTCGtA srTyGCCAAA gSGCCTgkks TGGG.ATGTT TACGCTGGTT



5  
10  
15  
20

```

251  GCCGkACTGC  CCGGCGTGT  TCTGTTCCGGC  TTTCCCGCAC  AGTTCATCAA
301  CGGCACGATT  AATTTCGTGGT  TCGGCAACGA  TACCCACGAG  GCGCTTGAAC
351  GCAGCCTCAA  TTTGAGCAAG  TCCGCATTGA  ATTTGGCGGC  AGACAACGCC
401  CTCGGCAACG  CCGTCCCGGT  GCAGATAGAC  CTCATCGGCG  CGGCTTCCCT
451  GCCCggggAT  ATGGGCAGGG  TGCTGGAACA  TTACGCCGGC  AGCGGTTTTG
501  CCCAGCTTGC  CCTGTACAA  ksCGCAAGCG  GCAAAATCGA  AAAAAGCATC
551  AACCCGCACA  AGCTCGATCA  GCCGTTTCCA  GGTAAGGCGC  GTTGGGAaAa
601  AATCCaACGG  GCGGGTTCGG  TCAGGGATTT  GGAAAGCATA  GGCGGCGTAT
651  TGTaCGCGCA  GGGCTGGCTG  TCGGCGGGTA  CGCACwACGG  GCGCGATTAC
701  GCCTTGT TTT  TCCGTCAGCC  GGTTCCCAAA  GGCGTGGCAG  AGGATGCCGT
751  yTTAATCGAA  AAGGCAAGGG  CGAAATATGC  TGAGTTGAGT  TACAGCAAAA
801  AAGGTTTGCA  GACCTTTTTC  CTGGCAACCC  TGCTGATTGC  CTCGCTGCTG
851  TCGATTTTTC  TTGCACTGGT  CATGGCACTG  TATTTCGCCC  GCCGTTTCCG
901  CGAACCCGTC  CTATCGCTTG  CCGAGGGGGC  GAAGGCGGTG  GCGCAAGGCG
951  ATTTAGCCA  GACGCGCCCG  GTGTTGCGCA  ACGACGAGTT  CGGACGCTTG
1001  ACCArGTTGT  TCAACCACAT  GACCGAGCAG  CTTTCCATCG  CCAAAGATGC
1051  AGACGAGCGC  AACC GCCGGC  GCGAGGAAGC  CGCCAGGCAT  TATCTTGAAT
1101  GCGTGTTGGA  GGGGCTGACC  ACGGGCGTGG  TGGTGTTTGA  CGAACAGGCG
1151  TGTCTGAAA  CCTTCAACAA  AGCGGCGGGT  ACC...

```

20 This corresponds to the amino acid sequence <SEQ ID 250; ORF64>:

25

```

1  MRRFLPIAAI  CAXXLXXGLT  AATGSTSSLA  DYFWWIVAFS  AMLLLVLSAV
51  LARYVILLLK  DRRDGVFGSX  XAKXPXXMF  TLVAXLPGVF  LFGFPAQFIN
101  GTINSWFGND  THEALERSLN  LSKSALNLAA  DNALGNAVVP  QIDLIGAASL
151  PGDMGRVLEH  YAGSGFAQLA  LYNXASGKIE  KSINPHKLDQ  PFPKGARWEK
201  IQRAGSVRDL  ESIGGVLYAQ  GWLSAGTHXG  RDYALFFRQP  VPKGVAEDAV
251  LIEKARAKYA  ELSYSKKGLQ  TFFLATLLIA  SLLSIFLALV  MALYFARFV
301  EPVLSLAEGA  KAVAQGDfsQ  TRPVLRNDEF  GRLTXLFNHM  TEQLSIAKDA
351  DERNRRREEA  ARHYLECVLE  GLTTGVVVD  EQGCLKTFNK  AAGT...

```

Further work revealed the complete nucleotide sequence <SEQ ID 251>:

30  
35  
40  
45  
50  
55  
60  
65

```

1  ATGCGCCGTT  TTCTACCGAT  CGCAGCCATA  TGCGCCGTCG  TCCTGTTGTA
51  CGGACTGACG  GCGGCAACCG  GCAGCACCAG  TTCGCTGGCG  GATTATTTCT
101  GGTGGATTGT  TGCGTTCAGC  GCAATGCTGC  TGCTGGTGT  GTCCGCCGTT
151  TTGGCACGTT  ATGTATATTT  GCTGTTGAAA  GACAGGCGCG  ACGGCGTATT
201  CGGTTTCGAG  ATTGCCAAAC  GCCTTTCTGG  GATGTTTACG  CTGGTTGCCG
251  TACTGCCCGG  CGTGTTTCTG  TTCGGCGTTT  CCGCACAGTT  CATCAACGGC
301  ACGATTAATT  CGTGGTTCGG  CAACGATACC  CACGAGGCGC  TTGAACGCAG
351  CCTCAATTG  AGCAAGTCCG  CATTGAATTT  GGCGGCAGAC  AACGCCCTCG
401  GGAACGCCGT  CCCCCTGCAG  ATAGACCTCA  TCGGCGCGGC  TTCCTGCCG
451  GGGGATATGG  GCAGGGTGCT  GGAACATTAC  GCCGCGAGCG  GTTTTGCCCA
501  GCTTGCCCTG  TACAATGCCG  CAAGCGGCAA  AATCGAAAAA  AGCATCAACC
551  CGCACAAGCT  CGATCAGCCG  TTTCCAGGTA  AGGCGCGTTG  GGAAAAAATC
601  CAACGGGCGG  GTTCGGTCAG  GGATTTGGAA  AGCATAGGCG  GCGTATTGTA
651  CGCGCAGGGC  TGGCTGTCCG  CGGGTACGCA  CAACGGGCGC  GATTACGCCT
701  TGTTTTTCCG  TCAGCCGGTT  CCCAAAGGCG  TGGCAGAGGA  TGCCGTCTTA
751  ATCGAAAAGG  CAAGGGCGAA  ATATGCTGAG  TTGAGTTACA  GCAAAAAGG
801  TTTGCAGACC  TTTTTCCTGG  CAACCCTGCT  GATTGCCTCG  CTGCTGTCGA
851  TTTTTCTTGC  ACTGGTCATG  GCACTGTATT  TCGCCCGCCG  TTTTCGTCGA
901  CCCGTCCCTAT  CGCTTGCCGA  GGGGGCGAAG  GCGGTGGCGC  AAGGCGATTT
951  CAGCCAGACG  CGCCCGTGT  TGCGCAACGA  CGAGTTCGGA  CGCTTGACCA
1001  AGTTGTTCAA  CCACATGACC  GAGCAGCTTT  CCATCGCCAA  AGAAGCAGAC
1051  GAGCGCAACC  GCCGCGCGCA  GGAAGCCGCC  AGGCATTATC  TTGAATGCGT
1101  GTTGGAGGGG  CTGACCACGG  GCGTGGTGGT  GTTTGACGAA  CAAGGCTGTC
1151  TGAAAACCTT  CAACAAGCG  GCGGAACAGA  TTTTGGGGAT  GCCGCTTACC
1201  CCCCTGTGGG  GCAGCAGCCG  GCACGGTTGG  CACGGCGTTT  CGGCGCAGCA
1251  GTCCCTGCTT  GCCGAAGTGT  TTGCCGCCAT  CGGCGCGGCG  GCAGGTACGG
1301  ACAAAACGGT  CCATGTGAAA  TATGCCGCGC  CGGACGATGC  CAAAATCCTG
1351  CTGGGCAAGG  CAACCGTCTT  GCCGAAGAC  AACGGCAACG  GCGTGTTAAT
1401  GGTGATTGAC  GACATCACCG  TTTTGATACA  CGCGCAAAAA  GAAGCCGCGT
1451  GGGGCGAAGT  GGCGAAGCGG  CTGGCACACG  AAATCCGCAA  TCCGCTCACG
1501  CCCATCCAGC  TTTCCGCCGA  ACGGCTGGCG  TGAAATTGG  GCGGGAAGCT
1551  GATGAGCAG  GATGCGCAAA  TCCTGACGCG  TTCGACCGAC  ACCATCGTCA
1601  AACAGGTGGC  GGCATTGAAG  GAAATGGTCG  AAGCATTCCG  CAATTATGCG
1651  CGTTCCTCTT  CGCTCAAATT  GGAAAATCAG  GATTTGAACG  CCTTAATCGG
1701  CGATGTGTTG  GCATTGTATG  AAGCCGTTCC  GTGCCGTTT  GCGGCGGAGC
1751  TTGCCGGCGA  ACCGCTGACG  GTGGCGGCGG  ATACGACCGC  CATGCGGCAG
1801  GTGCTGCACA  ATATTTTCAA  AAATGCCGCC  GAAGCGGCGG  AAGAAGCCGA

```

5  
1851 TGTGCCCGAA GTCAGGGTAA AATCGGAAAC AGGGCAGGAC GGTCGGATTG  
1901 TCCTGACGGT TTGCGACAAC GGCAAAGGGT TCGGCAGGGA AATGCTGCAC  
1951 AACGCCTTCG AGCCGTATGT AACGGACAAA CCGGCGGGAA CGGGATTGGG  
2001 TCTGCCTGTG GTGAAAAAAA TCATTGAAGA ACACGGCGGC CGCATCAGCC  
2051 TGAGCAATCA GGATGCGGGT GGCGCGTGTG TCAGAATCAT CTTGCCAAAA  
2101 ACGGTAAAAA CTTATGCGTA G

This corresponds to the amino acid sequence <SEQ ID 252; ORF64-1>:

10  
15  
20  
1 MRRFLPIAAI CAVVLLYGLT AATGSTSSLA DYFWWIVAFS AMLLLVLSAV  
51 LARYVILLK DRRDGVFGSQ IAKRLSGMFT LVAVLPGVFL FGVSAQFING  
101 TINSWFGNDT HEALERSLNL SKSALNLAAD NALGNAVFPVQ IDLIGAASLP  
151 GDMGRVLEHY AGSGFAQLAL YNAASGKIEK SINPHKLDQP FPGKARWEKI  
201 QRAGSVRDLE SIGGVLYAQQ WLSAGTHNGR DYALFFRQPV PKGVAEDAVL  
251 IEKARAKYAE LSYSKKGLQT FFLATLLIAS LLSIFLALVM ALYFARRFVE  
301 PVLSLAEGAK AVAQGDFSQT RPVLRNDEFG RLTKLFNHMT EQLSIAKEAD  
351 ERNRRREEAA RHYLECVLEG LTTGVVVFE QGCLKTFNKA AEQILGMLPT  
401 PLWGSSRHGW HGVSAQSSLL AEFVFAAIGAA AGTDKPVHVK YAAPDDAKIL  
451 LGKATVLPED NGNGVVMVID DITVLIHAQK EAAWGEVAKR LAHEIRNPLT  
501 PIQLSAERLA WKLGGKLDEQ DAQILTRSTD TIVKQVAALK EMVEAFRNYA  
551 RSPSLKLENQ DLNALIGDVL ALYEAGPCRF AAELAGEPLT VAADTTAMRQ  
601 VLHNIKFNAE EAAEEADVPE VRVKSETGQD GRIVLTVCDN GKGFGREMLH  
651 NAFEPYVTDK PAGTGLGLPV VKKIIIEHGG RISLSNQDAG GACVRIILPK  
701 TVKTYA\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

25 ORF64 shows 92.6% identity over a 392aa overlap with an ORF (ORF64a) from strain A of *N. meningitidis*:

30  
35  
40  
45  
50  
55  
60

orf64.pep	10	20	30	40	50	60
	MRRFLPIAAICAXLXGLTAATGSTSSLADYFWWIVAFSAMLVLSAVLARYVILLK					
orf64a	MRRFLPIAAICAVVLLYGLTAATGSTSSLADYFWWIVAFSAMLVLSAVLARYVILLK					
orf64.pep	70	80	90	100	110	120
	DRRDGVFGSXXXAKXPXXMFTLVAXLPGVFLFGFPAQFINGTINSWFGNDTHEALERSLN					
orf64a	DRRDGVFGSQIAKR-LSGMFTLVAVLPGVFLFGVSAQFINGTINSWFGNDTHEALERSLN					
orf64.pep	130	140	150	160	170	180
	LSKSALNLAADNALGNAVFPVQIDLIGAASLPDGMGRVLEHYAGSGFAQLALYNXASGKIE					
orf64a	LSKSALNLAADNALGNAIPVQIDXIGAASLPXDMGRVLEHYAGSGFAQLALYNAASGKIE					
orf64.pep	190	200	210	220	230	240
	KSINPHKLDQFPFGKARWEKIQRAGSVRDLESIGGVLYAQQWLSAGTHXGRDYALFFRQP					
orf64a	KSINPHKLDQFPFGKARWEKIQAGSVRDXSIGGVLYAXGWSAXTHNGRDYALFFRQP					
orf64.pep	250	260	270	280	290	300
	VPKGVAEDAVLIEKARAKYAE LSYSKKGLQTFFLATLLIASLLSIFLALVMALYFARRFV					
orf64a	VPKGVAEDAVLIEKARAXXXLSYSKKGLQTFFLATLLIASLLSIFLALVMALYFARRFV					
orf64.pep	310	320	330	340	350	360
	EPVLSLAEGAKAVAQGDFSQTRPVLRNDEFGRLTXLFNHMTQELSIKADADERNRRREEA					
orf64a	EPVLSLAEGAKAVAQGDFSQTRPVLRNDEFGRLTKLFNHMTQELSIKAEADERNRRREEA					

370 380 390  
orf64.pep ARHYLECVLEGLTTGVVVFDEQGCLKTFNKAAGT  
|||||  
orf64a ARHYLECVLEGLTTGVVVFDEQGCLKTFNKAEEQILGMPLTPLWGSSRHGHWGVSAAQQSL  
5 360 370 380 390 400 410  
orf64a LAEVFAAIGAAAGTDKPVHVKYAAPDDAKILLGKATVLPEDNXNGVVMVIDDITVLIHAQ  
420 430 440 450 460 470

The complete length ORF64a nucleotide sequence <SEQ ID 253> is:

10 1 ATGCGCCGTT TTCTACCGAT CGCAGCCATA TCGCGCCGTCG TCCTGTTGTA  
51 CCGACTGACG GCGGCAACCG GCAGCACCAG TTCGCTGGCG GATTATTTCT  
101 GGTGGATTGT TGCCTTCAGC GCAATGCTGC TGCTGGTGTT GTCCGCCGTT  
15 151 TTGGCAGGTT ATGTCATATT GCTGTTGAAA GACAGGCGCG ACGGCGTATT  
201 CGGTTTCGAG ATTGCGAAAC GCCTTTCCGG GATGTTTACG CTGGTTGCCG  
251 TACTGCCCGG CGTGTCTCTG TTCGGCGTTC CCGCACAGTT TATCAACGGC  
301 ACGATTAATT CGTGGTTCGG CAACGATACC CACGAGGCGC TTGAACGCAG  
351 CCTCAATTG AGCAAGTCCG CATTGAATCT GGCGGCAGAC AACGCCCTTG  
401 GCAACGCCAT CCCCGTCGAG ATAGACNTCA TCGGCGCGGC TTCCCTGCCC  
451 NGGGATATGG GCAGGTGCT GGAACATTAC GCCGCGAGCG GTTTTGCCCA  
20 501 GCTTGCCCTG TACAATGCCG CAAGCGGCAA AATCGAAAAA AGCATCAACC  
551 CGCACAGCT CGATCAGCCG TTCCAGGTA AGGCGCGTTG GGAAAAAATC  
601 CAACAGGCGG GTTCGGTCAG GGATNNGGAA AGCATAGGCG GCGTATTGTA  
651 CGCGCANGGC TGGCTGTCGG CAGNNACGCA CAACGGGCGC GATTACGCCT  
701 TGTTTTTCCG TCAGCCGGTT CCCAAAGGCG TGGCAGAGGA TGCCGCTCTA  
25 751 ATCGAAAAGG CAAGGGCGNA ANANNNTNAG TTGAGTTACA GCAAAAAAGG  
801 TTTGCAGACC TTTTCTCTNG CAACCCTGCT GATTGCCTCN CTGCTGTGCA  
851 TTTTCTTGC ACTGTCATG GCACTGTATT TCGCCCGCCG TTTCTGTCGAA  
901 CCCGTCCTAT CGCTTGCCGA GGGGGCGAAG GCGGTGGCGC AAGGCGATTT  
951 CAGCCAGACG CGCCCGTGT TCGCAACGA CGAGTTCGGA CGCTTGACCA  
30 1001 AGTTGTTCAA CCACATGACC GAGCAGCTTT CCATCGCCAA AGAAGCAGAC  
1051 GAGCGCAACC GCCGCGCGCA GGAAGCCGCC AGACATTATC TCGAATGCGT  
1101 GTTGGAGGGG CTGACCACGG GCGTGGTGGT GTTTGACGAA CAAGGCTGTC  
1151 TGAAAACTT CAACAAAGCG GCGGAACAGA TTTTGGGGAT GCCGCTTACC  
1201 CCCCTGTGGG GCAGCAGCCG GCACGGTTGG CACGGCGTTT CGGCGCAGCA  
35 1251 GTCCCTGCTT GCCGAAGTGT TTGCCGCCAT CGGCGCGCGG GCAGGTACGG  
1301 ACAACCCGTT CCATGTGAAA TATGCCGCGC CGGACGATGC CAAATCCTG  
1351 CTGGGCAAGG CAACGTCCT GCCGGAAGAC AACNGCAAC GCGTGGTAAT  
1401 GGTGATTGAC GACATCACCG TTTTGATACA CGCGCAAAAA GAAGCCGCGT  
1451 GGGGCGAAGT GGCAAAACGG CTGGCACACG AAATCCGCAA TCCGCTCACG  
40 1501 CCCATCCAGC TTTCTGCCGA ACGGCTGGCG TGGAATTGG GCGGGAAGCT  
1551 GGACGAGCAN GACGCGCAAA TCCTGACAG TTCGACCGAC ACCATCATCA  
1601 AACAAGTGGC GGCATTAAAA GAAATGGTCG AGGCATTCCG CAATTACNCG  
1651 CGTTCCCTTT CGNCTCAATT GGAAAATCAG GATTGTAACG CCTTAATCGG  
45 1701 CGATGTGTG GCATTGTACG AAGCTGGTCC GTGCCGTTT GCGGCGGAAC  
1751 TTGCCGCGCA ACCGCTGATG ATGGCGGCGG ATACGACCGC CATGCGGCAG  
1801 GTGCTGCACA ATATTTTCAA AAATGCCGCC GAAGCGCGCG AAGAAGCCGA  
1851 TGTGCCCGAA GTCAGGGTAA AATCGGAAGC GGGGCAGGAC GGACGGATTG  
1901 TCCTGACAGT TTGCGACAAC GGCAAGGGGT TCGGCAGGGA AATGCTGCAC  
1951 AATGCCCTTCG AGCCGTATGT AACGGACAAA CCGGCTGGAA CGGGATTGNG  
50 2001 ACTGCCCGTG GTGAAAAAAA TCATTGAAGA ACACGGCGGC CNCATCAGCC  
2051 TGAGCAATCA GGATGCGGGC GCGCGTNTG TCAGAATCAT CTTGCCAAAA  
2101 ACGGTAGAAA CTTATGCGTA G

This encodes a protein having amino acid sequence <SEQ ID 254>:

55 1 MRRFLPIAAI CAVVLLYGLT AATGSTSSLA DYFWWIVAFS AMLLLVLSAV  
51 LARYVILLK DRRDGVFGSQ IAKRLSGMFT LVAVLPGVFL FGVSAQFING  
101 TINSWFGNDT HEALERSLNL SKSALNLAAD NALGNAIPVQ IDXIGAASLP  
151 XDMGRVLEHY AGSGFAQLAL YNAASGKIEK SINPHKLDQP FPGKARWEKI  
201 QQAGSVRDXE SIGGVLYAXG WLSAXTHNGR DYALFFRQPV PKGVDAVL  
60 251 IEKARAXXXX LSYSKKGLQT FFLATLLIAS LLSIFLALVM ALYFARRFVE  
301 PVLSLAEGAK AVAQGDFSQT RPLRNDEFG RLTKLFNHMT EQLSIAKEAD  
351 ERNRRREEAA RHYLECVLEG LTTGVVVFDE QGCLKTFNKA AEQILGMPLT  
401 PLWGSSRHGW HGVSAQQSLL AEVFAAIGAA AGTDKPVHVK YAAPDDAKIL  
451 LGKATVLPED NXNGVVMVID DITVLIHAQK EAAWGEVAKR LAHEIRNPLT  
501 PIQLSAERLA WKLGGKLDX DAQILTRSTD TIIKQVAALK EMVEAFRNYX  
65 551 RSPSXQLENQ DLNALIGDVL ALYEAGPCR F AAELAGEPLM MAADTTAMRQ  
601 VLHNI FKNAA EAAEEADVPE VRVKSEAGQD GRIVLTVCDN GKGFGREMLH

651 NAFEPYVTDK PAGTGLXLPV VKKIIEHGG XISLSNQDAG GAXVRIILPK  
701 TVETYA\*

ORF64a and ORF64-1 show 96.6% identity in 706 aa overlap:

5 orf64a.pep 10 20 30 40 50 60  
MRRFLPIAAICAVVLLYGLTAATGSTSSLADYFWWIVAFSAMLVLSAVLARYVILLK  
orf64-1 MRRFLPIAAICAVVLLYGLTAATGSTSSLADYFWWIVAFSAMLVLSAVLARYVILLK  
10 orf64a.pep 70 80 90 100 110 120  
DRRDGVFGSQIAKRLSGMFTLVAVLPGVFLFGVSAQFINGTINSWFGNDTHEALERSNL  
orf64-1 DRRDGVFGSQIAKRLSGMFTLVAVLPGVFLFGVSAQFINGTINSWFGNDTHEALERSNL  
15 orf64a.pep 130 140 150 160 170 180  
SKSALNLAADNALGNAIPVQIDXIGAASLPXDMGRVLEHYAGSGFAQLALYNAASGKIEK  
orf64-1 SKSALNLAADNALGNAVFPVQIDLIGAASLPXDMGRVLEHYAGSGFAQLALYNAASGKIEK  
20 orf64a.pep 190 200 210 220 230 240  
SINPHKLDQFPFGKARWEKIQAGSVRDSEIGGVLYAGWLSAXTHNGRDYALFFRQPV  
orf64-1 SINPHKLDQFPFGKARWEKIQAGSVRDSEIGGVLYAGWLSAXTHNGRDYALFFRQPV  
25 orf64a.pep 250 260 270 280 290 300  
PKGVAEDAVLIEKARAXXXLSYSKKGLQTFLLIASLLSIFLALVMALYFARRFE  
orf64-1 PKGVAEDAVLIEKARAKYAELSYSKKGLQTFLLIASLLSIFLALVMALYFARRFE  
30 orf64a.pep 310 320 330 340 350 360  
PVLSLAEGAKAVAQGDFSQTRPVLNRNDEFGRITKLFNHMTQLSIAKEADERNRREEAA  
orf64-1 PVLSLAEGAKAVAQGDFSQTRPVLNRNDEFGRITKLFNHMTQLSIAKEADERNRREEAA  
35 orf64a.pep 370 380 390 400 410 420  
RHYLECVLEGLTTGVVVFDEQGCLKTFNKAAEQILGMPLTPLWGSSRHGWHGVSAQQSLL  
orf64-1 RHYLECVLEGLTTGVVVFDEQGCLKTFNKAAEQILGMPLTPLWGSSRHGWHGVSAQQSLL  
40 orf64a.pep 430 440 450 460 470 480  
AEVFAAIGAAAGTDKPVHVKYAAPDDAKILLGKATVLPEDNXNGVVMVIDDITVLIHAQK  
orf64-1 AEVFAAIGAAAGTDKPVHVKYAAPDDAKILLGKATVLPEDNGNGVVMVIDDITVLIHAQK  
45 orf64a.pep 490 500 510 520 530 540  
EAAWGEVAKRLAHEIRNPLTPIQLSAERLAWKLGGKLDQDAQILTRSTDTIKQVAALK  
orf64-1 EAAWGEVAKRLAHEIRNPLTPIQLSAERLAWKLGGKLDQDAQILTRSTDTIVKQVAALK  
50 orf64a.pep 550 560 570 580 590 600  
EMVEAFRNYXRSXQLENQDLNALIGDVLALYEAGPCRFAELAGEPLMAADTTAMRQ  
orf64-1 EMVEAFRNYARSPLKLENQDLNALIGDVLALYEAGPCRFAELAGEPLTVAADTTAMRQ  
55 orf64a.pep 610 620 630 640 650 660  
VLHNIFFKNAAEAAEADVPEVRVKSEAGQDGRIVLTVCDNGKGFGREMLHNAFEPYVTDK  
orf64-1 VLHNIFFKNAAEAAEADVPEVRVKSETQDGRIVLTVCDNGKGFGREMLHNAFEPYVTDK  
60 orf64a.pep 610 620 630 640 650 660  
VLHNIFFKNAAEAAEADVPEVRVKSEAGQDGRIVLTVCDNGKGFGREMLHNAFEPYVTDK  
orf64-1 VLHNIFFKNAAEAAEADVPEVRVKSETQDGRIVLTVCDNGKGFGREMLHNAFEPYVTDK  
65 orf64a.pep 610 620 630 640 650 660  
VLHNIFFKNAAEAAEADVPEVRVKSEAGQDGRIVLTVCDNGKGFGREMLHNAFEPYVTDK  
orf64-1 VLHNIFFKNAAEAAEADVPEVRVKSETQDGRIVLTVCDNGKGFGREMLHNAFEPYVTDK

		670	680	690	700
	orf64a.pep	PAGTGLXLPVVKKIIEEHGGXISLSNQDAGGAXVRILPKTVETYAX			
5	orf64-1	PAGTGLGLPVVKKIIEEHGGGRISLSNQDAGGACVRILPKTVKTYAX			
		670	680	690	700

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF64 shows 86.6% identity over a 387aa overlap with a predicted ORF (ORF64.ng) from *N.*

10 *gonorrhoeae*:

	orf64.pep	MRRFLPIAAICAXLXXGLTAATGSTSSLADYFWWIVAFSAMPLLLVLSAVLARYVILLK	60
	orf64ng	MRRFLPIAAICAVVLLYGLTAATGSTSSLADYFWWIVSFSAMPLLLVLSAVLARYVILLK	60
15	orf64.pep	DRRDGVFGSXXAKXPXXMFTLVAXLPGVFLFGFPAQFINGTINSWFGNDTHEALERSLN	120
	orf64ng	DRRNGVFGSQIAKR-LSGMFTLVAVLPGLFLFGISAQFINGTINSWFGNDTHEALERSLN	119
20	orf64.pep	LSKSALNLAADNALGNVAVPQIDLIGAASLPDGMGRVLEHYAGSGFAQLALYNXASGKIE	180
	orf64ng	LSKSALDLAADNAVSNVAVPQIDLIGTASLSGNMGSVLEHYAGSGFAQLALYNAASGKIE	179
	orf64.pep	KSINPHKLDQPPFGKARWEKIQRAGSVRDLESIGGVLYAQGWLSAGTHXGRDYALFFRQP	240
25	orf64ng	KSINPHQFDQPLDKEHWEQIQQTGSVRSLESIGGVLYAQGWLSAGTHNGRDYALFFRQP	239
	orf64.pep	VPKGVAEDAVLIEKARAKYAELSYSKKGLQTFFLATLLIASLLSIFLALVMALYFARRFV	300
	orf64ng	IPENVAQDAVLIEKARAKYAELSYSKKGLQTFFLVTLIASLLSIFLALVMALYFARRFV	299
30	orf64.pep	EPVLSLAEGAKAVAQGDQFSQTRPVLNRNDEFGRLTXLFNHMTQLSIAKDADERNRRREEA	360
	orf64ng	EPILSLAEGAKAVAQGDQFSQTRPVLNRNDEFGRLTKLFNHMTQLSIAKEADERNRRREEA	359
35	orf64.pep	ARHYLECVLEGLTTGVVVFDEQGCLKTFNKAAGT	394
	orf64ng	ARHYLECVLDGLTTGVVVSYP LSCCRTAVFSTCHSSPLSYF	400

An ORF64ng nucleotide sequence <SEQ ID 255> was predicted to encode a protein having amino acid sequence <SEQ ID 256>:

40	1	MRRFLPIAAI	CAVVLLYGLT	AATGSTSSLA	DYFWWIVSFS	AMLLLVLSAV
	51	LARYVILLK	DRRNGVFGSQ	IAKRLSGMFT	LVAVLPGLFL	FGISAQFING
	101	TINSWFGNDT	HEALERSLNL	SKSALDLAAD	NAVSNVAVPQ	IDLIGTASLS
	151	GNMGSVLEHY	AGSGFAQLAL	YNAASGKIEK	SINPHQFDQP	LPDKEHWEQI
45	201	QQTGSVRSLE	SIGGVLYAQG	WLSAGTHNGR	DYALFFRQPI	PENVAQDAVL
	251	IEKARAKYAE	LSYSKKGLQT	FFLVTLIAS	LLSIFLALVM	ALYFARRFVE
	301	PILSLAEGAK	AVAQGDQFSQ	RPVLRNDEFG	RLTKLFNHMT	EQLSIAKEAD
	351	ERNRRREEAA	RHYLECVLDG	LTGVVVSYP	LSCCRTAVFS	TCHSSPLSYF*

Further work revealed the complete gonococcal DNA sequence <SEQ ID 257>:

50	1	ATGCGCCGCT	TCCTACCGAT	CGCAGCCATA	TGCGCCGTCG	TCCTGCTGTA
	51	CGGATTGACG	GCGGCGACCG	GCAGCACCAG	TTCGCTGGCG	GATTATTTCT
	101	GGTGGATAGT	CTCGTTCAGC	GCAATGCTGC	TGCTGGTGTT	GTCCGCCGTT
	151	TTGGCACGTT	ATGTCATATT	GCTGTTGAAA	GACAGGCGCA	ACGGCGTGTT
	201	CGGTTCCGAG	ATTGCCAAAC	GCCTTTCCGG	GATGTTACAG	CTGGTCGCCG
	251	TACTGCCCGG	CTTGTTCCCTG	TTCGGCATTT	CCGCGCAGTT	TATCAACGGC
55	301	ACGATTAAAT	CGTGGTTCGG	CAACGACACC	CACGAAGCCC	TCAACGCAG
	351	CCTTAATTTG	AGCAAGTCCG	CACTGGATTT	GGCGGCAGAC	AATGCCGTCA
	401	GCAACGCCGT	TCCCGTACAG	ATAGACCTCA	TGGGCACCGC	CTCCCTGTCT
	451	GGCAATATGG	GCAGTGTGCT	GGAACACTAC	GCCGGCAGCG	GTTTGTGCCC
	501	GCTTGCCCTG	TACAATGCCG	CAAGCGGGAA	AATCGAAAAA	AGCATCAATC
60	551	CGCACCAATT	CGACCAGCCG	CTTCCCGACA	AAGAACATTG	GGAACAGATT

5  
10  
15  
20  
25  
30

```

601 CAGCAGACCG GTTCGGTTCG GAGTTTGGAA AGCATAGGCG GCGTATTGTA
651 CGCGCAGGGA TGGTTGTCGG CAGGTACGCA CAACGGGCGC GATTACGCGC
701 TGTTCCTCCG CCAGCCGATT CCCGAAAATG TGGCACAGGA TGCCGTTCTG
751 ATTGAAAAGG CGCGGGCGAA ATATGCCGAA TTGAGTTACA GCAAAAAGG
801 TTTGCAGACC TTTTCTCTGG TAACCTGCT GATTGCTCG CTGCTGTCGA
851 TTTTCTCTGC GCTGGTAATG GCACTGTATT TTGCCCGCCG TTTCTGCGAA
901 CCCATTCTGT CGCTTGCCGA GGGCGCAAAG GCGGTGGCGC AGGGTGATTT
951 CAGCCAGACG CGCCCCGTAT TGCGCAACGA CGAGTTCGGA CGTTTGACCA
1001 AGCTGTTCAA CCATATGACC GAGCAGCTTT CCATCGCCAA AGAAGCAGAC
1051 GAACGCAACC GCCGGCGCGA GGAAGCCGCC CGTCACTACC TCGAGTCCGT
1101 GTTGGATGGG TTGACTACCG GTGTGGTGGT GTTTGACGAA AAAGGCCGTT
1151 TGAACACCTT CAACAAGGCG GCGGAACAGA TTTTGGGGAT GCCGCTCGCC
1201 CCCCTGTGGG GCAGCAGCCG GCACGGTTGG CACGGCGTTT CGGCGCAGCA
1251 GTCCCTGCTT GCCGAAGTGT TtgcgcacAT CGGTGCGCGC GCAGGTACGG
1301 ACAACCCGGT CCAGGTGGAA TATGCCGCGC CGGACGATGC CAAAATCCTG
1351 CTGGGCAAGG CGACGGTATT GCCCGAAGAC AACGGCAACG GCGTGGTGAT
1401 GGTGATTGAC GACATCACCG TGCTGATACG CGCGCAAAAA GAAGCCGCGT
1451 GGGGTGAAGT GCGGAAGCGG CTGGCACACG AAATCCGCAA TCCGCTCAGC
1501 CCCATCCAGC TTTCCGCCGA ACGGCTGGCG TGGAATATGG GCGGGAAGCT
1551 GGACGATCAG GACGCGCAAA TCCTGACGCG TtcgACCGAC ACCATCATCA
1601 AACAGgtggc gCGTtAAAA GAAATGGTCG AGGCATTCCG CAATTACCGG
1651 CGCGCCCCCTT CGCTCAAACT GGAATATCAG GATTTGAACG CCTTAATCGG
1701 CGATGTTTTG GCCCTGTACG AAGCCGGCCC GTGCCGTTT GAGGCGGAAC
1751 TTGCCGGCGA ACCCGTGATG ATGGCGGCGG ATACGACCGC CATGCGGCAG
1801 GTGCTGCACA ATATTTTCAA AAATGCCGCC GAAGCGGCGG AAGAAGCCGA
1851 TATGCCCGAA GTCAGGGTAA AATCGGAAAC GGGGAGGAC GGACGGATTG
1901 TCCTGACGGT TTGCGACAAC GGCAAGGGAT TCGGCAAGGA AATGCTGCAC
1951 AATGCTTTTC AGCCGTATGT GACGGATAAG CCGGCGGGAA CGGGACTGGG
2001 TCTGCTGTGA GTGAAAAAAA TCATTGGAGA ACACGGCGGC CGCATCAGCC
2051 TGAGCAATCA GGATCGGGT GGGGCGTGTG TCAGAATCAT CTTGCCAAAA
2101 ACGGTAGAAA CTTATGCGTA G

```

This corresponds to the amino acid sequence <SEQ ID 258; ORF64ng-1>:

35  
40  
45

```

1 MRRFLPIAAI CAVVLLYGLT AATGSTSSLA DYFWWIVSFS AMLLLVL SAV
51 LARYVILLLK DRRNGVFGSQ IAKRLSGMFT LVAVLPGLFL FGISAQFING
101 TINSWFGNDT HEALERSLNL SKSALDLAAD NAVSNAVVPQ IDLIGTASLS
151 GNMGSVLEHY AGSGFAQLAL YNAASGKIEK SINPHQFDQP LPDKEHWEQI
201 QQTGSVRSLE SIGGVLYAQQ WLSAGTHNGR DYALFFRQPI PENVAQDAVL
251 IEKARAKYAE LSYSKGLQT FFLVTLLIAS LLSIFLALVM ALYFARREVE
301 PILSLAEGAK AVAQGDFSQT RPVLNRNDEF RLTKLNFHMT EQLSIAKEAD
351 ERNRRREEAA RHYLECVLDG LTTGVVVFDE KGRCLKFNKA AEQILGMPLA
401 PLWGSRRHGW HGVSAAQSL AEVFAAIGAA AGTDKPVQVE YAAPDDAKIL
451 LGKATVLPED NGNGVVMVID DITVLIRAQK EAAWGEVAKR LAHEIRNPLT
501 PIQLSAERLA WKLGKLDQD DAQILTRSTD TIIKQVAALK EMVEAFRNYA
551 RAPSLKLENQ DLNALIGDVL ALYEAGPCRF EAELAGEPLM MAADTTAMRQ
601 VLHNIKFNAE EAAEADMPE VVRKSETGQD GRIVLTVCDN KGFGKEMLH
651 NAFEPYVTDK PAGTGLGLPV VKKIIEGHGG RISLSNQDAG GACVRIILPK
701 TVETYA*

```

ORF64ng-1 and ORF64-1 show 93.8% identity in 706 aa overlap:

50  
55  
60  
65

```

orf64ng-1.pep 10 20 30 40 50 60
MRRFLPIAAICAVVLLYGLTAATGSTSSLA DYFWWIVSFSAMLLLVL SAVLARYVILLK
|||||:|||||
orf64-1 10 20 30 40 50 60
MRRFLPIAAICAVVLLYGLTAATGSTSSLA DYFWWIVAFSAML LVL SAVLARYVILLK

orf64ng-1.pep 70 80 90 100 110 120
DRRNGVFGSQIAKRLSGMFTLVAVLPGLFLFGISAQFINGTINSWFGNDTHEALERSLNL
|||:|||||:|||||:|||||:|||||
orf64-1 70 80 90 100 110 120
DRRDGVFGSQIAKRLSGMFTLVAVLPGLFLFGVSAQFINGTINSWFGNDTHEALERSLNL

orf64ng-1.pep 130 140 150 160 170 180
SKSALDLAADNAVSNVFPQIDLIGTASLSGNMGSVLEHYAGSGFAQLALYNAASGKIEK
|||||:|||||:|||||:|||||:|||||
orf64-1 130 140 150 160 170 180
SKSALNLAADNALGNVFPQIDLIGAASLPGMGRVLEHYAGSGFAQLALYNAASGKIEK

```

		190	200	210	220	230	240
	orf64ng-1.pep	SINPHQFDQPLDPKEHWEQIQQTGSVRSLESIGGVLYAQGWLSAGTHNGRDYALFFRQPI					
	orf64-1	SINPHKLDQPPGKARWEKIQRAGSVRDLESIGGVLYAQGWLSAGTHNGRDYALFFRQPV					
5		190	200	210	220	230	240
	orf64ng-1.pep	PENVAQDAVLIEKARAKYAELSYSKKGLQTFFLVTLIASLLSIFLALVMALYFARRFVE					
	orf64-1	PKGVAEDAVLIEKARAKYAELSYSKKGLQTFFLATLLIASLLSIFLALVMALYFARRFVE					
10		250	260	270	280	290	300
	orf64ng-1.pep	PENVAQDAVLIEKARAKYAELSYSKKGLQTFFLVTLIASLLSIFLALVMALYFARRFVE					
	orf64-1	PKGVAEDAVLIEKARAKYAELSYSKKGLQTFFLATLLIASLLSIFLALVMALYFARRFVE					
		250	260	270	280	290	300
	orf64ng-1.pep	PILSLAEGAKAVAQGDFSQTRPVLNRNDEFGRITKLFNHMTQLSIAKEADERNRRREEAA					
	orf64-1	PVLSLAEGAKAVAQGDFSQTRPVLNRNDEFGRITKLFNHMTQLSIAKEADERNRRREEAA					
15		310	320	330	340	350	360
	orf64ng-1.pep	PILSLAEGAKAVAQGDFSQTRPVLNRNDEFGRITKLFNHMTQLSIAKEADERNRRREEAA					
	orf64-1	PVLSLAEGAKAVAQGDFSQTRPVLNRNDEFGRITKLFNHMTQLSIAKEADERNRRREEAA					
		310	320	330	340	350	360
	orf64ng-1.pep	RHYLCVLDGLTTGVVVFDEKGRITKFNKAAEQILGMPLAPLWGSSRHGWHGVSQAQSL					
	orf64-1	RHYLCVLEGLTTGVVVFDEQGCITKFNKAAEQILGMPLTPLWGSSRHGWHGVSQAQSL					
20		370	380	390	400	410	420
	orf64ng-1.pep	RHYLCVLDGLTTGVVVFDEKGRITKFNKAAEQILGMPLAPLWGSSRHGWHGVSQAQSL					
	orf64-1	RHYLCVLEGLTTGVVVFDEQGCITKFNKAAEQILGMPLTPLWGSSRHGWHGVSQAQSL					
		370	380	390	400	410	420
25		430	440	450	460	470	480
	orf64ng-1.pep	AEVFAAIGAAAGTDKPVQVEYAAPDDAKILLGKATVLPEDNGNGVVMVIDDITVLIRAQK					
	orf64-1	AEVFAAIGAAAGTDKPVHVKYAAPDDAKILLGKATVLPEDNGNGVVMVIDDITVLIHAQK					
		430	440	450	460	470	480
30		490	500	510	520	530	540
	orf64ng-1.pep	EAAWGEVAKRLAHEIRNPLTPIQLSAERLAWKLGKLDQDAQILTRSTDTIHKQVAALK					
	orf64-1	EAAWGEVAKRLAHEIRNPLTPIQLSAERLAWKLGKLDQDAQILTRSTDTIVKQVAALK					
35		490	500	510	520	530	540
	orf64ng-1.pep	EMVEAFRNYARAPSLKLENQDLNALIGDVLALYEAGPCRFAELAGEPLMMAADTTAMRQ					
	orf64-1	EMVEAFRNYARAPSLKLENQDLNALIGDVLALYEAGPCRFAELAGEPLTVAADTTAMRQ					
40		550	560	570	580	590	600
	orf64ng-1.pep	EMVEAFRNYARAPSLKLENQDLNALIGDVLALYEAGPCRFAELAGEPLMMAADTTAMRQ					
	orf64-1	EMVEAFRNYARAPSLKLENQDLNALIGDVLALYEAGPCRFAELAGEPLTVAADTTAMRQ					
		550	560	570	580	590	600
	orf64ng-1.pep	VLHNI FKNAAEAAEADMPVVRVKSETGQDGRIVLTVCNKGFGKEMLHNAFEPYVTDK					
	orf64-1	VLHNI FKNAAEAAEADMPVVRVKSETGQDGRIVLTVCNKGFGREMLHNAFEPYVTDK					
45		610	620	630	640	650	660
	orf64ng-1.pep	VLHNI FKNAAEAAEADMPVVRVKSETGQDGRIVLTVCNKGFGKEMLHNAFEPYVTDK					
	orf64-1	VLHNI FKNAAEAAEADMPVVRVKSETGQDGRIVLTVCNKGFGREMLHNAFEPYVTDK					
		610	620	630	640	650	660
50		670	680	690	700		
	orf64ng-1.pep	PAGTGLGLPVVKKIIEHGGRIISLNQDAGGACVRIILPKTVETYAX					
	orf64-1	PAGTGLGLPVVKKIIEHGGRIISLNQDAGGACVRIILPKTVETYAX					
		670	680	690	700		

Furthermore, ORF64ng-1 shows significant homology to a protein from *A. caulinodans*:

55	sp Q04850 NTRY_AZOCA NITROGEN REGULATION PROTEIN NTRY >gi 77479 pir  S18624 ntrY protein - Azorhizobium caulinodans >gi 38737 (X63841) NtrY gene product [Azorhizobium caulinodans] Length = 771 Score = 218 bits (550), Expect = 7e-56 Identities = 195/720 (27%), Positives = 320/720 (44%), Gaps = 58/720 (8%)
60	Query: 7 IAAICAVVLLYGLTAATGSTSSLDYFWWIXXXXXXXXXXXXXXXXXRYVILLKDRRNGV 66 I+A+ ++L GLT + + + R + + K R G Sbjct: 35 ISALATFLILMGLTPVVPVTHQVVIS----VLLVNAAVLILSAMVGREIWRIAKARARG 90
65	Query: 67 FGSQIAKRLSGMFTLVAVLPGLFLFGISAQFINGTINSWFGNDTHEALERSLNLSKLSALD 126 +++ R+ G+F +V+V+P + + +++ ++ ++ WF T E + S+++++ + Sbjct: 91 AAARLHIRIVGLFAVVSVPAILVAVVASLTLDRLDRWFSMRTQEIVASSVSVAQTYVR 150

Query: 127 LAADNAVSNAPVQIDLIGTASLSGNMGSVLEHYAG--SGFAQLALYNAASGKIEKSINP 184  
 A N + + + DL S+ Y G S F Q+ AA + + +  
 Sbjct: 151 EHALNIRGDILAMSADLRLKSV-----YEGDRSRFNQILTAQAALRNLPGLAMLI 200

5 Query: 185 HQFDQPLPDKEHWEQIQQTGSVRSLESIGGVLYAQGWLSAGTHNGRDYA----- 233  
 + D + + + I + V + +IG Q + N DY  
 Sbjct: 201 RR-DLSVVERAN-VNIGREFIVPANLAIGDATPDQPVYILP--NDADYVAAVVPLKDYDD 256

10 Query: 234 --LFFRQPIPENVAQDAVLIEKARAKYAELSYSKGLQTFFLVTXXXXXXXXXXXXXVMA 291  
 L+ + I V ++ A Y L + G+Q F + +  
 Sbjct: 257 LYLYVARLIDPRVIGYLKTTQETLADYRSLEERRFGVQVAFALMYAVITLIVLLSAVWLG 316

15 Query: 292 LYFARRFVEPILSLAEGAKAVAQGDFSQTRPVLNRND-EFGRLTKLFNHMTQELSIXXXXX 350  
 L F++ V PI L A VA+G+ P+ R + + L + FN MT +L  
 Sbjct: 317 LNFSKWLVAPIRRLMSAADHVAEGLDVRVPIYRAEGDLASLAETFNKMTHELRSQREAI 376

20 Query: 351 XXXXXXXXXXXXHYLECVLDGLTTGVVVFDEKGRCLKTFNKAAEQILGMPLAPLWGSRRHW 410  
 + E VL G+ GV+ D + R+ N++AE++LG L+ + RH  
 Sbjct: 377 LTARDQIDSRRRFTEAVLSGVGAGVIGLDSQERITILNRSARLLG--LSEVEALHRHLA 434

25 Query: 468 VIDDITVLIRAQKEAAWGEVAKRLAHEIRNPLTPIQLSAERLAWKLGKLDQDAQILTR 527  
 +DDIT LI AQ+ +AW +VA+R+AHEI+NPLTPIQLSAERL K G + QD +I +  
 Sbjct: 489 TLDDITELISAQRTSAWADVARRIAHEIKNPLTPIQLSAERLKRKFGRHV-TQDREIFDQ 547

30 Query: 528 STDIIKQVAALKEMVEAFRNYARAPSLKLENQDLNALIGDVLALYEAGPCRFEAELAGE 587  
 TDTII+QV + MV+ F ++AR P +++QD++ +I + L G + +  
 Sbjct: 548 CTDIIIRQVGDIGRMVDEFSSFARMFKPVVDSQDMSEIIRQTVFLMRVGHPEVVFDSEVP 607

35 Query: 588 PLMMAA-DTTAMRQVLHNIKFNXXXXXXXXDMPEVRVK-----SETGQDGRIVLTVCD 639  
 P M A D + Q L NI KN P+VR + + G+D +V+ + D  
 Sbjct: 608 PAMPARFDRRLVSQALTNILKNAEAIEAVP-PDVRGQGRIRVSANRVGED--LVIDIID 664

40 Query: 640 NGKGFGKEMLHNAFEPYVTDKPGTGLGLPVVKIIGEHGGRISLSNQDAG-GACVRIIL 698  
 NG G +E + EPYVT + GTGLGL +V KI+ EHGG I L++ G GA +R+ L  
 Sbjct: 665 NGTGLPQESRNRLLEPYVTTREKGTGLGLAIVGKIMEEHGGGIELNDAPEGRGAWIRLTL 724

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### 45 Example 31

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 259>:

1 ATGTACGCAT TTACCGCCGC ACAGCAACAG AAGGCACTCT TCCGGCTGGT  
 51 GCTTTTTCAT ATCCTCATCA TCGCCGCCAG CAACTATCTG GTGCAGTTCC  
 101 CTTTCCAAAT TTTCGGCATC CACACCACTT GGGGCGCATT TTCCTTTCCC  
 50 151 TTCATCTTCC TTGCCACCGA CCTGACCGTC CGCATTTTCG GTTCTCACTT  
 201 GGCACGGCGG ATTATCTTTT GGGTGATGTT CCCCGCCCTT TTGCTTTTCT  
 251 ACGTCTTTTC CGTTTGTGTC CACAACGGCA GTTGACAGG CTTGGGCGCG  
 301 CTGTCCGAAT TCAACACCTT TGTGGGACGC ATCGCCTTAG CCAGCTTTCC  
 351 CGCCTACGCG ATCGGACAAA TCCTTGATAT TTTGTATTAC AACAAATTAC  
 55 401 GCCGTCTGAA AGCGTGGTGG ATTGCACCGA ACGCATCAAC CGTCATCGGG  
 451 CACGCGTTGG ATACG...

This corresponds to the amino acid sequence <SEQ ID 260; ORF66>:

1 MYAFTAAQQQ KALFRLVLFH ILIIAASNYL VQFPFQIFGI HTTWGAFSFP



5	1	ATGTACGCAT	TTACCGCCGC	ACAGCAACAG	AAGGCACTCT	TCCGGCTGGT
	51	GCTTTTTTCAT	ATCCTCATCA	TCGCCGCCAG	CAACTATCTG	GTGCAAGTTT
	101	CTTTCCAAAT	TTTCGGCATC	CACACCACCT	GGGGCGCATG	TTCTTTTCCC
	151	TTCATCTTTC	TTGCCACCGA	CCTGACCGTC	CGGATTTTCG	TTCTCACTT
	201	GGCACGGCGG	ATTATCTTTT	GGGTGATGTT	CCCCGCCCTT	TTGCTTTTCT
10	251	ACGTCTTTTC	CTTTTGTGTC	CACAACGGCA	GTTGGACAGG	CTTGGGCGCG
	301	CTGTCCGAAT	TGAACACCTT	TGTCGGACGC	ATCGCCCTTAG	CCAGCTTTGC
	351	CGCCTACGCG	ATCGGACAAA	TCCTTGATAT	TTTTGTATTC	AACAAATTAC
	401	GCGCTCTGAA	AGCGTGGTGG	ATTGCACCGA	CCGCATCAAC	CGTCATCGGG
	451	AACGCCTTGG	ATACGGCTGT	ATTTTTCGCC	GTTGCTCTCT	ACGCAAGCAG
15	501	CGATGGATT	ATGGCGGCAA	ACTGGCAGGG	CATCGCTTTT	STCGATTACC
	551	TGTTCAAAT	TACCGTCTGC	ACCCTCTTCT	TCCTGCCCGC	CTACGGCGTG
	601	ATACTGAATC	TGCTGACGAA	AAAATGACA	ACCTTGCAAA	CCAAACAGGC
	651	GCAAGACCGC	CCCGCGCCCT	CGCTGCAAAA	TCCGTAA	

20

1	MYAFTAAQQQ	KALFRVLVLFH	ILIIAASNYL	VQFPFOIFGI	HTTWGAFSFP
51	FIFLATDLTV	RIFGSHLARR	IIFWVMFPAL	LLSVFVSFLV	HNGSWTGLGA
101	LSEFNTFVGR	IALASPAAYA	IGQILDIFVF	NKLRRLKAWW	IAPTASTVIG
151	NALDTLVFFA	VAFYASSDGF	MAANWQGIAT	VDYLFKLTV	TLFFLPAYGV
201	ILNLLTKKLT	TLQTKQAQDR	PAPSLQNP*		

Homology with the hypothetical protein o221 of *E. coli* (accession number P37619)

30	orf66	1	MYAFTAAQQQKALFRLVLFHILIIAASNVLVQFPFQIFGIHTTWGAFSFPFIFLATDLTV	60
			M F+ Q+ KALF L LFH+L+I +SNVLVQ P I G HTTWGAFSFPFIFLATDLTV	
	o221	1	MNVFSQTQRYKALFWLSLFLHLLVITSSNVLVQLPVSILGFHTTWGAFSFPFIFLATDLTV	60
35	orf66	61	RIFGSHLARRIIFWVMFPALLSYVFSVLFHNGSWTGLGALSEFNTEFVGRIALASFAAYA	120
			RIFG+ LARRIIF VM PALL+SYV S LF+ GSW G GAL+ FN FV RIA ASF AYA	
	o221	61	RIFGAPLARRIIFAVMIPALLISYVISSLFYMGSWQGFALAHFNLFVARIATASF MAYA	120
	orf66	121	IGQILDIFVFNKLRLKAWWIAPNASTVIGHALDT	155
			+GQILD+ VFN+LR+ + WW+AP AST+ G+ DT	
	o221	121	LGQILDVHVFENRLRQSRRWLAPTASTLFGNVSDT	155

ORF66 shows 96.1% identity over a 155aa overlap with an ORF (ORF66a) from strain A of *N. meningitidis*:

```

45      orf66.pep      MYAFTAAQQQKALFRLVLFHILIIAASNYLVQFPFQIFGIHTTWGAFSFPFI10 20 30 40 50 60FLATDLTV
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      orf66a          MYAFTAAQQQKALFWLVLFHILIIAASNYLVQFPFQISGIHTTWGAFSFPFI10 20 30 40 50 60FLATDLTV
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

50      orf66.pep      RIFGSHLARRIIFWVMFPALLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA70 80 90 100 110 120
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      orf66a          RIFGSHLARRIIFWVMFPALLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA70 80 90 100 110 120
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

55      orf66.pep      IGQILDIFVFNKLRLKAWWIAPNASTVIGHALDT130 140 150
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

**orf66a**      LGQILDIFVFNKLRRLKAWWVAPTASTVIGNALDTLVFFAVAFYASSDGFMAANWQGIAF

                130                140                150                160                170                180

orf66a      VDYLFKLTVCGLFFLPAYGVILNLLTKKLTTLQTKQAQDRPAPSLQNPX  
                    190            200            210            220

The complete length ORF66a nucleotide sequence <SEQ ID 263> is:

	1	ATGTACGCAT	TTACCGCCGC	ACAGCAACAG	AAGGCACCT	TCTGGCTGGT
	51	GCTTTTTCAT	ATCCTCATCA	TCGCCGCCAG	CAACTATCTG	GTGCAGTTCC
	101	CCTTCCAAAT	TTCCGGCATC	CACACCACCT	GGGGCGCGTT	TTCTTTTCCC
10	151	TTCATCTTCC	TCCGCCACGA	CCTGACCGTC	CGCATTTTCG	GTTTCGCACT
	201	GGCACGGCGG	ATTATCTTTT	GGTTCATGTT	CCCCGCCCTT	TTGCTTTTCT
	251	ACGTCTTTTC	CGTTTTGTTC	CACAACGGCA	GTTGGACGGG	CTTGGGCGCG
	301	CTGTCCGAAT	TCAACACCTT	TGTCGGACGC	ATCGCGCTGG	CAAGTTTTCG
	351	CGCCTACGCG	CTCGGACAAA	TCCTTGATAT	TTTTGTGTTT	AACAAATTAC
15	401	GCCGTCTGAA	AGCGTGGTGG	GTTGCCCCGA	CTGATCAAC	CGTCACTGGC
	451	AACGCCTTAG	ATACGTTGGT	ATTTTTCGCC	GTTGCCTTCT	ACGCAAGCAG
	501	CGATGGATTT	ATCGCGCAAA	ACTGGCAGGG	CATCGCTTTT	GTGCATTACC
	551	TGTTCAAAC	CACCGTCTCG	GGTCTGTTTT	TCTGCCCCG	CTACGGCGTG
	601	ATTCTGAATC	TGCTGACGAA	AAACTGTGAC	ACCTTGCAAA	CCAAACAGGC
20	651	GCAAGACGCG	CCCGCGCCCT	CGCTGCAAAA	TCCGTA	

This encodes a protein having amino acid sequence <SEQ ID 264>:

25

1	MYAFTAAQQQ	KALFWLVLFH	ILIIAASNYL	VQFPFOISGI	HTTWGAFSFP
51	FIFLATDLTV	RIFGSHLARR	IFFWMFPPAL	LLSYVFSVLF	HNGSWTGLGA
101	LSEFTFVGR	IALASFAAYA	LGQILDIVF	NKLRRLKAWW	VAPTASTVIG
151	NALDTLVFFA	VFYASSDGF	MAANWQGIAF	VDYLFKLTYC	GLFFLPAYGV
201	ILNLLTKKLT	TLQTKQAQDR	PAPSLQNP*		

ORF66a and ORF66-1 show 97.8% identity in 228 aa overlap:

		10	20	30	40	50	60
30	orf66a.pep	MYAFTAAQQQKALFWLVL	FHILIIAASN	YLVPFPFQISGIHTTWGAFS	FPFFIFLATDLTV		
	orf66-1	MYAFTAAQQQKALFRLVL	FHILIIAASN	YLVPFPFQIFGIHTTWGAFS	FPFFIFLATDLTV		
		10	20	30	40	50	60
		70	80	90	100	110	120
35	orf66a.pep	RIFGSHLARRIIFWVMFP	ALLSYVFSVL	FHNGSWTGLG	ALSEFNTFVGRI	ALASFAAYA	
	orf66-1	RIFGSHLARRIIFWVMFP	ALLSYVFSVL	FHNGSWTGLG	ALSEFNTFVGRI	ALASFAAYA	
		70	80	90	100	110	120
40		130	140	150	160	170	180
	orf66a.pep	LGQILDIFVFNKLRLKAW	WVAPTASTV	IGNALDTLVFF	AVAFYASSDGF	MAANWQGI	AF
		:	:	:	:	:	:
	orf66-1	IGQILDIFVFNKLRLKAW	WIAPTASTV	IGNALDTLVFF	AVAFYASSDGF	MAANWQGI	AF
		130	140	150	160	170	180
45							
		190	200	210	220	229	
	orf66a.pep	VDYLFKLTVCGLF	FLPAYGVILN	LLTKKLTTLQ	TKQAQDRPAP	SLQNPK	
	orf66-1	VDYLFKLTVC	TFLFLPAYGV	ILNLLTKKLT	TTLQTKQAQDR	PAPSLQNPK	
50		190	200	210	220		

Homology with a predicted ORF from *N.gonorrhoeae*

ORF66 shows 94.2% identity over a 155aa overlap with a predicted ORF (ORF66.ng) from *N. gonorrhoeae*:

[illegible]

```

orf66ng      RIFGSHLARRIIFWVMFPALSLSYVFSVLFHNGSWTGLGAPSQFNTFVGRIALASFAAYA 120
orf66.pep    IGQILDIFVFNKLRLKAWWIAPNASTVIGHALDT 155
              :|||||:|||||:|||||:|||||:|||||
5 orf66ng    LGQILDIFVFDKLRRLKAWWIAPAASTVIGNALDTLVFFAVAFYASSDEFMAANWQGI AF 180

```

The complete length ORF66ng nucleotide sequence <SEQ ID 265> is:

```

1 ATGTACGCAT TGACCGCCGC ACAGCAACAG AAGGCACTCT TCCGGCTGGT
51 GCTTTTCCAT ATCCTCATCA TCGCCGCCAG CAACTATCTG GTGCAGTCC
101 CTTTCCGGAT TTTCCGCATC CACACCACTT GGGGCGCGTT TTCCTTTCCC
10 151 TTCATCTTCC TCGCCACCGA CCTGACCGTC CGCATTTTCG GTTCGCACTT
201 GCGCGCGCGG ATTATCTTTT GGGTGATGTT CCCC GCCCCTT ttgCTTTcat
251 aCGTCTTTTC CGTTTGTTC CACAACGGCA GTTGGACGGG CTTGGGCGCG
301 ctgTCCCAAT TCAACACCTT TGTCCGACGC ATCGCGCTGG CAAGTTTTCG
351 CGCCTACGCG CTCGGACAAA TCCTTGATAT TTTCTGATTG GACAAATTAC
15 401 GCCGTCTGAA AGCGTGGTGG ATTGCCCCGG CCGCATCAAC CGTCATCGGC
451 AATGCACTGG ACACGTTAGT ATTTTTCGCC GTTGCCTTTT ACGCAAGCAG
501 CGATGAATTT ATGGCGGCAA ACTGGCAGGG CATCGCTTTT GTCGATTACC
551 TGTTCAAATC TACCGTCTGC ACCCTCTTCT TCCTGCCCGC CTACGGCGTG
601 ATACTGAATC TGCTGACGAA AAAACTGACG GCCCTGCAAA CCAAACAGGC
20 651 GCAAGACCGC CCCGTGCCCT CGCTGCAAAA TCCGTAA

```

This encodes a protein having amino acid sequence <SEQ ID 266>:

```

1 MYALTAAQQQ KALFRLVLFH ILIIAASNYL VQFPFRIFGI HTTWGAFSFP
51 FIFLATDLTV RIFGSHLARR IIFWVMFPAL SLSYVFSVLF HNGSWTGLGA
101 PSQFNTFVGR IALASFAAYA LGQILDIFVF DKLRRLKAWW IAPAASTVIG
25 151 NALDTLVFFA VAFYASSDEF MAANWQGI AF VDYLFLKLTVC TLFFLPAYGV
201 ILNLLTKKLT ALQTKQAQDR PVPSLQNP*

```

An alternative annotated sequence is:

```

1 MYALTAAQQQ KALFRLVLFH ILIIAASNYL VQFPFRIFGI HTTWGAFSFP
51 FIFLATDLTV RIFGSHLARR IIFWVMFPAL SLSYVFSVLF HNGSWTGLGA
30 101 LSQFNTFVGR IALASFAAYA LGQILDIFVF DKLRRLKAWW IAPAASTVIG
151 NALDTLVFFA VAFYASSDEF MAANWQGI AF VDYLFLKLTVC TLFFLPAYGV
201 ILNLLTKKLT ALQTKQAQDR PVPSLQNP*

```

ORF66ng and ORF66-1 show 96.1% identity in 228 aa overlap:

```

orf66-1.pep MYAFTAAQQQKALFRLVLFHILIIAASNYLVQFPFQIFGIHTTWGAFSFPFIFLATDLTV 60
35 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf66ng      MYALTAAQQQKALFRLVLFHILIIAASNYLVQFPFRIFGIHTTWGAFSFPFIFLATDLTV 60

orf66-1.pep RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA 120
40 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf66ng      RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSQFNTFVGRIALASFAAYA 120

orf66-1.pep IGQILDIFVFNKLRLKAWWIAPTASTVIGNALDTLVFFAVAFYASSDGFMANWQGI AF 180
45 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf66ng      LGQILDIFVFDKLRRLKAWWIAPAASTVIGNALDTLVFFAVAFYASSDEFMAANWQGI AF 180

orf66-1.pep VDYLFLKLTVCTLFFLPAYGVILNLLTKKLTTLQTKQAQDRPAPSLQNPX 229
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf66ng      VDYLFLKLTVCTLFFLPAYGVILNLLTKKLTALQTKQAQDRPVPSLQNPX 229

```

Furthermore, ORF66ng shows significant homology with an *E.coli* ORF:

```

50 sp|P37619|YHHQ_ECOLI HYPOTHETICAL 25.3 KD PROTEIN IN FTSY-NIKA INTERGENIC
   REGION (O221)
   >gi|1073495|pir||S47690 hypothetical protein o221 - Escherichia coli >gi|466607
   (U00039) No definition line found [Escherichia coli] >gi|1789882 (AE000423)
55 hypothetical 25.3 kD protein in ftsY-nikA intergenic region [Escherichia coli]
   Length = 221
   Score = 273 bits (692), Expect = 5e-73
   Identities = 132/203 (65%), Positives = 155/203 (76%)

Query: 1 MYALTAAQQQKALFRLVLFHILIIAASNYLVQFPFRIFGIHTTWGAFSFPFIFLATDLTV 60

```

M + Q+ KALF L LFH+L+I +SNYLVO P I G HTTWGAFSFPFIFLATDLTV  
 Sbjct: 1 MNVFSQTQRYKALFWLSLFHLLVITSSNYLVQLPVSILGFHTTWGAFSFPFIFLATDLTV 60  
 Query: 61 RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSQFNTFVGRIALASFAAYA 120  
 RIFG+ LARRIIF VM PALL+SYV S LF+ GSW G GAL+ FN FV RIA ASF AYA  
 Sbjct: 61 RIFGAPLARRIIFAVMIPALLISYVISSLFYMGSWQGFALAHFNLFVARIATASEMAYA 120  
 Query: 121 LGQILDIEFVFDKLRRLKAWWIAPAASTVIGNALDTLVFFAVAFYASSDEEMAANWQGI AF 180  
 LGQILD+ VF++LR+ + WW+AP AST+ GN DTL FF +AF+ S D FMA +W IA  
 Sbjct: 121 LGQILDVHVFNRLRQSRWWLAPTASTLFGNVSDTLAFFFFIAFWRSPDAFMAEHWMEIAL 180  
 Query: 181 VDYLKLTVCTLFPLPAYGVILN 203  
 VDY FK+ + +FFLP YGV+LN  
 Sbjct: 181 VDYCFKVLISIVFFLPMYGVLLN 203

Based on this analysis, including the homology with the *E.coli* protein and the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

## Example 32

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 267>:

1 ATGGTCATAA AATATACAAA TTTGAATTTT GCGAAATTGT CGATAATTGC  
 51 AATTTTGATG ATGTATTCGT TTGAAGCGAA TGCAAATGCA GTmwrAATAT  
 101 CTGAAACTGT TTCAGTTGAT ACCGGACAAG GTGCGAAAAT TCATAAGTTT  
 151 GTACCTAAAA ATAGTAAAAC TTATTCATCT GATTTAATAA AAACGGTAGA  
 201 TTAAACACAC AyyCCTACGG GCGCAAAAGC CCGAATCAAC GCCAAAATAA  
 251 CCGCCAGCGT ATCCCGCGCC GCGGTATTGG CCGGGGTCGG CAAACTTGCC  
 301 CGCTTAGGCG CGAAATTCAG CACAAGGGCG GTtCCCTATG TCGGAACAGC  
 351 CcTTTtagcc CACGACGTAT ACGAAAcTTT CAAAGAAGAC ATACAGGCAC  
 401 GAGGCTACCA ATACGACCCC GAAACCGACA AATTTGTAA AGGCTACGAA  
 451 TATAGTAATT GCCTTTGGTA CGAAGACAAA AGACGTATTA ATAGAACCTA  
 501 TGGCTGCTAC GCGTTGAT.

This corresponds to the amino acid sequence <SEQ ID 268; ORF72>:

1 MVIKYTNLNF AKLSIIAILM MYSFEANANA VXISETVSVD TGQGAHIKHF  
 51 VPKNSKTYSS DLIKTVDLTH XPTGAKARIN AKITASVSRA GVLAGVVKLA  
 101 RLGAKFSTRA VPYVGTALLA HDVYETFKED IQARGYQYDP ETDKFEVKGYE  
 151 YSNCLWYEDK RRINRTYGCY GVD..

Further work revealed the complete nucleotide sequence <SEQ ID 269>:

1 ATGGTCATAA AATATACAAA TTTGAATTTT GCGAAATTGT CGATAATTGC  
 51 AATTTTGATG ATGTATTCGT TTGAAGCGAA TGCAAATGCA GTAAAAATAT  
 101 CTGAAACTGT TTCAGTTGAT ACCGGACAAG GTGCGAAAAT TCATAAGTTT  
 151 GTACCTAAAA ATAGTAAAAC TTATTCATCT GATTTAATAA AAACGGTAGA  
 201 TTAAACACAC ATCCCTACGG GCGCAAAAGC CCGAATCAAC GCCAAAATAA  
 251 CCGCCAGCGT ATCCCGCGCC GCGGTATTGG CCGGGGTCGG CAAACTTGCC  
 301 CGCTTAGGCG CGAAATTCAG CACAAGGGCG GTTCCCTATG TCGGAACAGC  
 351 CCTTTtagcc CACGACGTAT ACGAAAcTTT CAAAGAAGAC ATACAGGCAC  
 401 GAGGCTACCA ATACGACCCC GAAACCGACA AATTTGCAAA GGTCTCAGGC  
 451 TAA

This corresponds to the amino acid sequence <SEQ ID 270; ORF72-1>:

1 MVIKYTNLNF AKLSIIAILM MYSFEANANA VKISETVSVD TGQGAHIKHF  
 51 VPKNSKTYSS DLIKTVDLTH IPTGAKARIN AKITASVSRA GVLAGVVKLA  
 101 RLGAKFSTRA VPYVGTALLA HDVYETFKED IQARGYQYDP ETDKFAKVS  
 151 \*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF72 shows 98.0% identity over a 147aa overlap with an ORF (ORF72a) from strain A of *N.*

*meningitidis*:

```

5      10      20      30      40      50      60
orf72.pep  MVIKYTNLNF AKLSIIAILM MYSFEANANAVK ISETVSVDTGQGAKIHKFVPKNSKTYSS
orf72a     MVIKYTNLNF AKLSIIAILM MYSFEANANAVK ISETVSVDTGQGAKIHKFVPKNSKTYSS
10      70      80      90      100     110     120
orf72.pep  DLIKTVDLTHXPTGAKARINAKITASVSRAGVLAVGK LARLGAKFSTRVPPYVGTALLA
orf72a     DLIKTVDLTHIPTGAKARINAKITASVSRAGVLAVGK LARLGAKFSTRVPPYVGTALLA
15      70      80      90      100     110     120
orf72.pep  HDVYETFKEDIQARGYQYDPETDKFVKGYEYSNCLWYEDKRRINRITYGCGYVD
orf72a     HDVYETFKEDIQARGYQYDPETDKFAKVS GX
20      130     140     150

```

The complete length ORF72a nucleotide sequence <SEQ ID 271> is:

```

1  ATGGTCATAA AATATACAAA TTTGAATTTT GCGAAATTGT CGATAATTGC
51 AATTTTGATG ATGTATTCGT TTGAAGCGAA TGCAAATGCA GTAAAAATAT
151 CTGAAACTGT TTCAGTTGAT ACCGGACAAG GTGCGAAAAT TCATAAGTTT
201 TTTAACACAC ATCCCTACGG GCGCAAAAGC CCGAATCAAC GCCAAAATAA
251 CCGCCAGCGT ATCCCGCGCC GCGGTATTGG CCGGGGTCGG CAAACTTGCC
301 CGCTTAGGCG CGAAATTCAG CACAAGGGCG GTTCCCTATG TCGGAACAGC
351 CCTTTAGGCC CACGACGTAT ACGAACTTT CAAAGAAGAC ATACAGGCAC
401 GAGGCTACCA ATACGACCCC GAAACCGACA AATTTGCAAA GGTCTCAGGC
451 TAA

```

This encodes a protein having amino acid sequence <SEQ ID 272>:

```

1  MVIKYTNLNF AKLSIIAILM MYSFEANANA VKISETVSVDTGQGAKIHKF
51 VPKNSKTYSS DLIKTVDLTH IPTGAKARIN AKITASVSRA GVLAVGKLA
101 RLGAKFSTRA VPPYVGTALLA HDVYETFKED IQARGYQYDP ETDKFAKVS
151 *

```

ORF72a and ORF72-1 show 100.0% identity in 150 aa overlap:

```

40      10      20      30      40      50      60
orf72a.pep  MVIKYTNLNF AKLSIIAILM MYSFEANANAVK ISETVSVDTGQGAKIHKFVPKNSKTYSS
orf72-1     MVIKYTNLNF AKLSIIAILM MYSFEANANAVK ISETVSVDTGQGAKIHKFVPKNSKTYSS
45      70      80      90      100     110     120
orf72a.pep  DLIKTVDLTHIPTGAKARINAKITASVSRAGVLAVGK LARLGAKFSTRVPPYVGTALLA
orf72-1     DLIKTVDLTHIPTGAKARINAKITASVSRAGVLAVGK LARLGAKFSTRVPPYVGTALLA
50      70      80      90      100     110     120
orf72a.pep  HDVYETFKEDIQARGYQYDPETDKFAKVS GX
orf72-1     HDVYETFKEDIQARGYQYDPETDKFAKVS GX
55      130     140     150

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF72 shows 89% identity over a 173aa overlap with a predicted ORF (ORF72.ng) from *N.*

*gonorrhoeae*:

```

5      orf72.pep      MVIKYTNLNF AKLSIIAILM MYSFEANANAVKISETVSVDTGQGAKIHKFVPKNSKTYSS      60
      orf72ng      MVTKHTNLF AKLSIIAILM MYSFEANANAVKISETVSVDTGQGAKVHKFVPKSSNIYSS      60
      orf72.pep      DLIKTVDLTHXPTGAKARINAKITASVSRAGVLAVGKGLARLGAKFSTRAPVYVGTALLA      120
10     orf72ng      DLTKAVDLTHIPTGAKARINAKITASVSRAGVLSGVGKLVQRQAKFGTRAPVYVGTALLA      120
      orf72.pep      HDVYETFKEDIQARGYQYDPETDKFVKGYEYSNCLWYEDKRRINRTYGCYGVD      173
      orf72ng      HDVYETFKEDIQARGCRYDPETDKFVKGYEYANCLWYEDERRINRTYGCYGVDDSSIMRLM      180

```

15 An ORF72ng nucleotide sequence <SEQ ID 273> was predicted to encode a protein having amino acid sequence <SEQ ID 274>:

```

      1  MVTKHTNLF AKLSIIAILM MYSFEANANA VKISETVSVD TGQGAKVHKF
51  VPKSSNIYSS DLTKAVDLTH IPTGAKARIN AKITASVSRAGVLSGVGKLV
101 RQGAKEGTRA VPYVGTALLA HDVYETFKED IQARGCRYDP ETDKFKVGYE
201 YANCLWYEDE RRINRTYGCY GVDSSIMRLM PDRSRFPEVK QLMESQMYRL
251 ARPFWNWRKE ELNKLSSLDW NNFVLNRCTF DWNGGGCAVN KGDDFRAGAS
251 FSLGRNPXYK EEMDAKKPEE ILSLKVDADP DKYIEATGYP GYSEKVEVAP
301 GTKVNMGPVT DRNGNPVQVA ATFGDRDAQN TTADVQVIPR PDLTPASAEA
351 PHAQPLPEVS PAENPANNPD PDENPGTREN PEPDPLNPD ANPDTGQPG
401 TSPDSPAVPD RPNGRHRKER KEGEDGGLSC DYFPEILACQ EMGKPSDRMF
451 HDISIPQVTD DKTWSSHNF PSNGVCPQPK TFHVFGQYR ASYEPLCVFA
501 EKIRFAVLLA FIIMSAFVVF GSLGGE*

```

After further analysis, the following gonococcal DNA sequence <SEQ ID 275> was identified:

```

      1  ATGGTCACAA AACATACAAA TTTGAATTTT GCGAAATGTG CGATAATTGC
51  AATTTTGATG ATGTATTCTG TTGAAGCGAA TGCAAATGCA GTAAAAATAT
101 CTGAAACTCT TTCGGTTGAT ACCGGACAAG GCGCGAAAGT TCATAAGTTC
151 GTTCCTAAAT CAAGTAATAT TTATTCATCT GATTTAACAA AAGCGGTAGA
201 TTTAACGCAT ATCCCGACGG GCGCAAAAGC CCGAATCAAC GCCAAAATAA
251 CCGCCAGCGT ATCCCGCGCC GCGGTATTGT CGGGGGTCGG CAAACTTGTC
301 CGCCAAGGCG CGAAATTCGG CACAAGGCG GTTCCTATG TCGGAACAGC
351 CCTTTTAGCC CACGACGTAT ACGAACTTT CAAAGAAGAC ATACAGGCAC
401 GAGGCTGCCG ATACGATCCC GAAACCGACA AATTT

```

This corresponds to the amino acid sequence <SEQ ID 276; ORF72ng-1>:

```

40      1  MVTKHTNLF AKLSIIAILM MYSFEANANA VKISETVSVD TGQGAKVHKF
      51  VPKSSNIYSS DLTKAVDLTH IPTGAKARIN AKITASVSRAGVLSGVGKLV
      101 RQGAKEGTRA VPYVGTALLA HDVYETFKED IQARGCRYDP ETDKFK

```

ORF72ng-1 and ORF721-1 show 89.7% identity in 145 aa overlap:

```

      10      20      30      40      50      60
45     orf72ng-1.pe MVTKHTNLF AKLSIIAILM MYSFEANANAVKISETVSVDTGQGAKVHKFVPKSSNIYSS
      orf72-1      MVIKYTNLNF AKLSIIAILM MYSFEANANAVKISETVSVDTGQGAKIHKFVPKNSKTYSS
      10      20      30      40      50      60

      70      80      90      100     110     120
50     orf72ng-1.pe DLTKAVDLTHIPTGAKARINAKITASVSRAGVLSGVGKLVQRQAKFGTRAPVYVGTALLA
      orf72-1      DLIKTVDLTHIPTGAKARINAKITASVSRAGVLAVGKGLARLGAKFSTRAPVYVGTALLA
      70      80      90      100     110     120

55     orf72ng-1.pe HDVYETFKEDIQARGCRYDPETDKF
      orf72-1      HDVYETFKEDIQARGCRYDPETDKF

```

### Example 33

10

1	ATGAGATTTT	TCGGTATCGG	TTTTTTGGTG	CTGCTGTTTT	TGGAGATTAT
51	GTCGATTGTG	TGGGTTGCCG	ATTGGCTGGG	CGGCGGCTGG	ACGTTGTTTT
101	TGATGGCGGC	AGGTTTTGCC	GCCGGCGTGC	TGATGCTCAG	GCAAACCGGG
151	<u>G</u> CTGACCGGT	CTTTTATTGG	CGGGCGCGGC	AATGAGAAGC	GGCGGGAAGG
201	TATCCGTTTA	TCAGATGTGG	TGGCCTATC..		

1 MRFFGIGFLV LLFLEIMSIV WVADWLGGGW TLFLMAAGFA AGVLMRLQTG  
51 LTGLLLAGAA MRSGGKVSFY QMLWPI..

20	1	ATGAGATTTT	TCGGTATCGG	TTTTTTGGTG	CTGCTGTTTT	TGGAGATTAT
	51	GTCGATTGTG	TGGGTTGCCG	ATTGGCTGGG	CGGCGGCTGG	ACGTTGTTTT
	101	TGATGGCGGC	AGGTTTGGCC	GCCGGCGTGC	TGATGCTCAG	GCATACGGGG
	151	CTGTGCCGTG	TTTTATTGGC	GGCGCGGCCA	ATGAGAAGCG	CGCGAGAGGGT
	201	ATCCGTTTAT	CAGATGTTGT	GGCCTATCCG	TTATACGGTG	CGCGCTGTGT
25	251	GTCTGATGAG	TCCGGGATTG	GTATCCTCGG	TGTTGGCGGT	ATTGCTGCTT
	301	TGCGCGTTTA	AGGGAGGGGC	AGTGTTCGAC	GCAGAGAGGT	CGAAAAATTT
	351	TTTCAACATG	AACCAATCGG	GCAGAAAAGA	GGGCTTTTCC	CGCGATGACG
	401	ATATTATCGA	GGGAGAATAT	ACGGTTGAAG	AGCCTTACGG	CGGCAATCGT
	451	TCCCGAAACG	CCATCGAACA	CAAAAAAGAC	GAAATAA	

30 1 MRFFGIGFLV LLFLEIMSVV WVADWLGGGW TLFMAAGFA AGVLMRLRHTG  
51 LSGLLLAGAA MRSGGRVSVY QMLWPIRYTV AAVCLMSPGF VSSVLAVLLL  
101 LPFKGGAVLQ AGGAENFFNM NQSGRKEGFS RDDDIIEGEY TVEEYPGGNR  
151 SRNAIEHKKD E\*

35 Homology with a predicted ORF from *N.meningitidis* (strain A)

		10	20	30	40	50	60
40	orf73.pep	<u>MRFFGIGIFLVLLFLEIMSI VVWVADWLGGGWTLFLMAAGFAAGVLMRQTGLTGILLLAGAA</u>					
		:     :     :					
	orf73a	<u>MRFFGIGIFLVLLFLEIMSI VVWVADWLGGGWTLFLMAATFAAGVVMRLRHTGLSLGILLLAGAA</u>					
		10	20	30	40	50	60
		70					
45	orf73.pep	MRSGGKVS VYQMLWPI					
		:					
	orf73a	MRSGGRVS VYXMLWXIRYTVAAVCXMSPGFVSSVXAVLLXLPFKGGAVLOAGGAENFFNM					

The complete length ORF73a nucleotide sequence <SEQ ID 281> is:

```

      1 ATGAGATTTT TCGGTATCGG TTTTGTGGTG CTGCTGTTTT TGGAGATTAT
      51 GTCGATTGTG TGGGTTGCCG ATTGGTTGGG CGGCGGTTGG ACGCTGTTTC
    101 TAATGGCGGC AACCTTTGCC GCCGCGTGG TGATGCTCAG GCATACGGGG
    151 CTGTCCGGTC TTTTATTGGC GGGCGCGGCA ATGAGAAGCG GCGGGAGGGT
    201 ATCCGTTTAT CANATGTTGT GGCNTATCCG TTATACGGTG GCGGCGGTGT
    251 GTCNGATGAG TCCGGGATTC GTATCCTCGG TGTNGCGCGT ATTGCTGN TG
    301 CTNCCGTTTA AGGGAGGTGC AGTGTTCAG GCAGGAGGTG CGGAAAATT
    351 TTTCAACATG AACCANTCGG GCAGAAAAGA NGGCNTTTC CGCGATGACG
    401 ATATTATCGA GGGGAATAT ACGGTTGAAG ANCCTTACGG CGGCANTCGT
    451 TTCCGAAACG CCNTNGAACA CAAAAAGAC GAATAA
  
```

This encodes a protein having amino acid sequence <SEQ ID 282>:

```

      1 MRFFGIGFLV LLFLEIMSIV WVADWLGGGW TFLMAATFA AGVVMRLHTG
      51 LSGLLLAGAA MRSGGRVSVY XMLWXIRYTV AAVCMSPGF VSSVXAVLLX
    101 LPFKGGAVLQ AGGAENFFNM NXSGRKXGXS RDDDIIEGEY TVEXPYGGXR
    151 FRNAXEHKKD E*
  
```

ORF73a and ORF73-1 show 91.3% identity in 161 aa overlap

```

      10      20      30      40      50      60
    20 orf73a.pep MRFFGIGFLVLLFLEIMSIVWVADWLGGGWTFLMAATFAAGVVMRLRHTGLSGLLLAGAA
        |||||:|||||
    20 orf73-1 MRFFGIGFLVLLFLEIMSIVWVADWLGGGWTFLMAAGFAAGVLMRLRHTGLSGLLLAGAA
        10      20      30      40      50      60

      70      80      90      100     110     120
    25 orf73a.pep MRSGGRVSVYXMLWXIRYTVAAVCMSPGFVSSVXAVLLXLPFKGGAVLQAGGAENFFNM
        |||||:|||||
    25 orf73-1 MRSGGRVSVYQMLWPIRYTVAAVCLMSPGFVSSVLAVLLLFPFKGGAVLQAGGAENFFNM
        70      80      90      100     110     120

      130     140     150     160
    30 orf73a.pep NXSGRKXGXS RDDDIIEGEYTVEXPYGGXRFRNAXEHKKDEX
        |||||:|||||
    30 orf73-1 NQSGRKEGFSRDDDIIEGEYTVVEPYGGNRSR NAEIHKKDEX
        130     140     150     160
  
```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF73 shows 92.1% identity over a 76aa overlap with a predicted ORF (ORF73.ng) from *N. gonorrhoeae*:

```

      40 orf73.pep MRFFGIGFLVLLFLEIMSIVWVADWLGGGWTFLMAAGFAAGVLMRLRQTGLTGLLLAGAA 60
        |||||:|||||
      40 orf73ng MRFFGIGFLVLLFLEIMSIVWVADWLGGGWTFLMAATFAAGVLMRLRHTGLSGLLLAGAA 60

      orf73.pep MRSGGKVSQMLWPI 76
        ::|||
    45 orf73ng VKSSGKVSQMLWPIRYTVAAVCLMSPGFVSSVLAVLLLFPFKGGAVLQAGGAENFFNM 120
  
```

The complete length ORF73ng nucleotide sequence <SEQ ID 283> is:

```

      1 ATGAGATTTT TCGGTATCGG TTTTGTGGTG CTGCTGTTTT TGAAATTAT
      51 GTCGATTGTG TGGGTTGCCG ATTGGCTGGG CGGCGGTTGG AcgcTGTTTC
    101 TAATGGCGGC AACCTTTGCC GCCGCTGTGC TGATGCTCAG GCATAcggGG
    151 CTGTCCGGTC TTTTATTGGC TGGCGCGGCG GTAAAAagta gtgGGAAGGT
    201 ATCTGTTTAT CagatgtTGT GGCCTATCCG TTATAcggtg gcggcggtgT
    251 GTCTGatgag tCgggGATTC GTATCCTccg tgttgCGGT ATTGCTGCTG
    301 CTGCcgttta aggGaggGgc agtgttgag gcaggagggtg cggaaaATT
    351 TTTCAACATg aaCcaatcgg gcagaaAaga gggatttttc cacgatgacg
    401 atattatcga gggagaatat acggttgaaa aacctgacgg cggcaatcgt
    451 tccgaAAcG ccatcgaaca cgaaaAagac gaataA
  
```

This encodes a protein having amino acid sequence <SEQ ID 284>:



```

1 MRFFGIGFLV LLFLEIMSIV WVADWLGGGW TLFLMAATFA AGVLMRLHTG
51 LSGLLLAGAA VKSSGKVSIV QMLWPIRYTV AAVCLMSPGF VSSVLAVLLL
101 LPEKGGAVLQ AGGAENFFNM NQSGRKEGFF HDDDIIEGEY TVEKPDGGR
151 SRNAIEHEKD E*

```

### 5 ORF73ng and ORG73-1 show 93.8% identity in 161 aa overlap

```

10 orf73-1.pep 10 20 30 40 50 60
    MRFFGIGFLVLLFLEIMSIVWVADWLGGGWTLFLMAAGFAAGVLMRLHTGLSGLLAGAA
    |||||
orf73ng      10 20 30 40 50 60
    MRFFGIGFLVLLFLEIMSIVWVADWLGGGWTLFLMAATFAAGVLMRLHTGLSGLLAGAA

15 orf73-1.pep 70 80 90 100 110 120
    MRSGRVSVYQMLWPIRYTVAAVCLMSPGFVSSVLAVLLLPEKGGAVLQAGGAENFFNM
    ::::|:|||||
orf73ng      70 80 90 100 110 120
    VKSSGKVSIVQMLWPIRYTVAAVCLMSPGFVSSVLAVLLLPEKGGAVLQAGGAENFFNM

20 orf73-1.pep 130 140 150 160
    NQSGRKEGFSRDDDIIEGEYTVVEPYGGRSRNAIEHKKDEX
    ||||| :|||||:| |||||:|
orf73ng      130 140 150 160
    NQSGRKEGFHDDDIIEGEYTVKPDGGRSRNAIEHEKDEX

```

Based on this analysis, including the presence of a putative leader sequence and putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 34

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 285>:

```

30 1 ATGTTTGT TFCAGACGGC ATTCTT.ATG TTTCAGAAAC ATTTGCAGAA
    51 AGCCTCCGAC AGCGTCGTCG GAGGGACATT ATACGTGGTT GCCACGCCCA
    101 TCGGCAATTT GCGCGACATT ACCCTGCGCG CTTTGGCGGT ATTGCAAAAG
    151 GCG..... .GCGA AGACACGCGC GTTACCGCAC AGCTTTTGAG
    201 CGCGTACGGC ATTCAGGGCA AACTCGTCAG TGTGCGCGAA CACAACGAAC
    251 GGCAGATGGC GGACAAGATT GTCGGCTATC TTTCAGACGG CATGGTTGTG
35 301 GCACAGGTTT CCGATGCGGG TACGCCGGCC GTGTGCGACC CGGGCGCGAA
    351 ACTCGCCCGC CGCGTGCCTG AGGCCGGGTT TAAAGTCGTT CCCGTCGTGG
    401 GCGCAAC.GC GGTGATGGCG GCTTTGAGCG TGGCCGGTGT GGAAGGATCC
    451 GATTTTATT TCAACGGTTT TGTACCGCCG AAATCGGGAG AACGCAGGAA
    501 ACTGTTTGCC AAATGGGTGC GGGCGGCGTT TCCTATCGTC ATGTTTGAAA
40 551 CGCGCACCG CATCGGTGCA GCGCTTGCCG ATATGGCGGA ACTGTTCCCC
    601 GAACGCCGAT TAATGCTGGC GCGCGAAATT ACGAAAACGT TTGAAACGTT
    651 CTTAAGCGGC ACGGTTGGGG AAATTCAGAC GGCATTGTCT GCCGACGGCG
    701 ACCAATCGCG CGGCGAGATG GTGTTGGTGC TTTATCCGGC GCAGGATGAA
45 751 AAACACGAAG GCTTGTCCGA GTCCGCGCAA AACATCATGA AAATCCTCAC
    801 AGCCGAGCTG CCGACCAAAC AGGCGGCGGA GCTTGCTGCC AAAATCACGG
    851 GCGAGGGAAA GAAAGCTTTG TACGAT..

```

This corresponds to the amino acid sequence <SEQ ID 286; ORF75>:

```

50 1 MFVFQTAFXM FQKHLQKASD SVVGGTLYVV ATPIGNLADI TLRALAVLQK
    51 A...AEDTR VTAQLLSAYG IQGKLVSVRE HNERQMADKI VGYLSDGMVV
    101 AQVSDAGTPA VCDPGAKLAR RVREAGFKVV PVVGAXAVMA ALSVAGVEGS
    151 DFYFNGFVFP KSGERRKLFA KVVRAAFPIV MFETPHRIGA ALADMAELFP
    201 ERRMLLAREI TKTFTFLSG TVGEIQTALS ADGDQSRGEM VLVLPAQDE
    251 KHEGLSESAQ NIMKILTAEI PTKQAELAA KITGEGKKAL YD..

```

Further work revealed the complete nucleotide sequence <SEQ ID 287>:

1 ATGTTTCAGA AACATTTGCA GAAAGCCTCC GACAGCGTCG TCGGAGGGAC  
 51 ATTATACGTG GTTGCCACGC CCATCGGCAA TTTGGCGGAC ATTACCCGTG  
 101 GCGCTTTGGC GGTATTGCAA AAGGCGGACA TCATCTGTGC CGAAGACACG  
 151 CGCGTTACCG CACAGCTTTT GAGCGGTAC GGCATTGAGG GCAAACTCGT  
 201 CAGTGTGCGC GAACACAACG AACGGCAGAT GGCGGACAAG ATTGTGCGGT  
 251 ATCTTTCAGA CGGCATGGTT GTGGCACAGG TTTCCGATGC GGGTACGCCG  
 301 GCCGTGTGCG ACCCGGGGCG GAAACTCGCC CGCCGCGTGC GTGAGGCCGG  
 351 GTTTAAAGTC GTTCCCGTCG TGGGCGCAAG CGCGGTGATG GCGGCTTTGA  
 401 GCGTGCCCGG TGTGGAAGGA TCCGATTTT ATTTCACGG TTTGTACCG  
 451 CCGAAATCGG GAGAACGCAG GAAACTGTT GCCAAATGGG TGCGGGCGCG  
 501 GTTTCCTATC GTCATGTTG AAACGCCGCA CGCATCGGT GCGACGCTTG  
 551 CCGATATGGC GGAAGTGTTC CCCGAACGCC GATTAATGCT GGCAGCGGAA  
 601 ATTACGAAAA CGTTTGAAAC GTTCTTAAGC GGCACGGTTG GGGAAATTCA  
 651 GACGGCATTG TCTGCCGACG GCAACCAATC GCGCGCGCAG ATGGTGTGCG  
 701 TGCTTTATCC GCGCGAGGAT GAAAAACACG AAGGCTGTG CAGTCCGCG  
 751 CAAAACATCA TGAAATCCT CACAGCCGAG CTGCCGACCA AACAGGCGGC  
 801 GGAGCTTGCT GCCAAAATCA CGGCGAGGG AAAGAAAGCT TTGTACGATC  
 851 TGGCTCTGTC TTGAAAAAC AATAG

This corresponds to the amino acid sequence <SEQ ID 288; ORF75-1>:

1 MFQKHLQKAS DSVVGGTLYV VATPIGNLAD ITRLALAVLQ KADIICAEDT  
 51 RVTAQLLSAY GIQGLVSVR EHNERQMDK IVGYLSDGMV VAQVSDAGTP  
 101 AVCDPGAKLA RRVREAGFKV VPVVGASAVM AALSVAAGVEG SDFYFNGFVP  
 151 PKSGERRKLF AKWVRAAFPI VMFETPHRIG ATLADMAELF PERRMLLARE  
 201 ITKTFETFLS GTVGEIQTAL SADGNQSRGE MVLVLYPAQD EKHEGLSESA  
 251 QNIMKILTAE LPTKQAAELA AKITGEGKKA LYDLALSWKN K\*

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF75 shows 95.8% identity over a 283aa overlap with an ORF (ORF75a) from strain A of *N. meningitidis*:

30		10	20	30	40	50	60
	orf75.pep	MFVFQTA	FXMFQKHLQKAS	DSVVGGTLYV	VATPIGNLADIT	LRALAVLQKAXXXA	EEDTR
	orf75a		MFQKHLQKAS	DSVVGGTLYV	VATPIGNLADIT	LRALAVLQKADIICA	EEDTR
35		70	80	90	100	110	120
	orf75.pep	VTAQLLSAY	GIQGLVSVREHNERQ	MADKIVGYLSDGMV	VAQVSDAGTP	AVCDPGAKLAR	
	orf75a	VTAQLLSAY	GIQGLVSVREHNERQ	MADKIVGYLSDGMV	VAQVSDAGTP	AVCDPGAKLAR	
40		60	70	80	90	100	110
	orf75.pep	RVREAGFKV	VPVVGASAVMAALS	VAGVEGSDFYFNGFV	PPKSGERRKLF	AKWVRAAFPIV	
	orf75a	RVREAGFKV	VPVVGASAVMAALS	VAGVEGSDFYFNGFV	PPKSGERRKLF	AKWVRAAFPIV	
45		120	130	140	150	160	170
	orf75.pep	MFETPHRIG	AALADMAELF	PERRMLLAREIT	KTFETFLSGTV	GEIQTALADGNQ	SRGEM
	orf75a	MFETPHRIG	AALADMAELF	PERRMLLAREIT	KTFETFLSGTV	GEIQTALADGNQ	SRGEM
50		180	190	200	210	220	230
	orf75.pep	VLVLYPAQ	DEKHEGLSESA	QNIMKILTAE	LPTKQAAELA	AKITGEGKKA	LYD
	orf75a	VLVLYPAQ	DEKHEGLSESA	QNIMKILTAE	LPTKQAAELA	AKITGEGKKA	LYDLALSWKNK
55		240	250	260	270	280	290
	orf75a	VLVLYPAQ	DEKHEGLSESA	QNIMKILTAE	LPTKQAAELA	AKITGEGKKA	LYDLALSWKNK
60	orf75a	X					

The complete length ORF75a nucleotide sequence <SEQ ID 289> is:

1 ATGTTTCAGA AACATTTGCA GAAAGCCTCC GACAGCGTCG TCGGAGGGAC  
 51 ATTATACGTG GTTGCCACGC CCATCGGCAA TTTGGCGGAC ATTACCCTGC  
 101 GCGCTTTGGC GGTATTGCAA AAGGCGGACA TCATCTGTGC CGAAGACACG  
 151 CGCGTTACCG CGCAGCTTTT GAGCGCGTAC GGCATTGAGG GCAAACCTCGT  
 201 CAGCGTGCGC GAACACAACG AACGGCAGAT GGCGGACAAG ATTGTGCGGT  
 251 ATCTTTTCAGA CGGCATGGTT GTGGCACAGG TTTCCGATGC GGGTACGCCG  
 301 GCCGTGTGCG ACCCGGGCGC GAAACTCGCC CGCCGCGTGC GTGAGGTCGG  
 351 GTTTAAAGTT GTCCTGTGTT TCGGCGCAAG CGCGGTGATG GCGGCTTTGA  
 401 GTGTGGCTGG TGTGGCGGGA TCCGATTTT ATTTCACGG TTTTGTACCG  
 451 CCGAAATCGG GCGAACGTAG GAAATTGTTT GCCAAATGGG TGCGGGTGGC  
 501 GTTTCCGTC GTGATGTTT AAACGCCGCA CGCATCGGG GCGACGCTTG  
 551 CCGATATGGC GGAAGTGTTC CCCGAACGCC GATTAATGCT GCGCGCGGAA  
 601 ATCACGAAAA CGTTTGAAAC GTTCTTAAGC GGCACGGTTG GGGAAATCA  
 651 GACGGCATTG GCGGCGGACG GCAACCAATC GCGCGCGGAG ATGGTGTG  
 701 TGCTTTATCC GCGCGAGGAT GAAAAACACG AAGGCTGTG CAGTCCGCG  
 751 CAAAACATCA TGAAATCCT CACAGCCGAG CTGCCGACCA AACAGGCGGC  
 801 GGAGCTTGCC GCCAAAATCA CGGCGAGGG AAAAAAGCT TTGTACGATC  
 851 TGGCACTGTC TTGAAAAAC AAATGA

This encodes a protein having amino acid sequence <SEQ ID 290>:

20 1 MFQKHLQKAS DSVVGGTLYV VATPIGNLAD ITRLALAVLQ KADIICAEDT  
 51 RVTAQLLSAY GIQKLVSVR EHNERQMDK IVGYLSDGMV VAQVSDAGTP  
 101 AVCDPGAKLA RRVREVGFKV VPVVGASAVM AALSVAGVAG SDFYFNGFVP  
 151 PKSGERRKLF AKWVRVAFPV VMFETPHRIG ATLADMAELF PERRMLLARE  
 201 ITKTFETFLS GTVGEIQTAL AADGNQSRGE MVLVLYPAQD EKHEGLSESA  
 251 QNIMKILTAE LPTKQAAELA AKITGEGKKA LYDLALSWKN K\*

ORF75a and ORF75-1 show 98.3% identity in 291 aa overlap:

		10	20	30	40	50	60
orf75a.pep		MFQKHLQKASDSVVGGTLYVVATPIGNLADITLRLALAVLQKADIICAEDTRVTAQLLSAY					
orf75-1		MFQKHLQKASDSVVGGTLYVVATPIGNLADITLRLALAVLQKADIICAEDTRVTAQLLSAY					
		10	20	30	40	50	60
orf75a.pep		GIQKLVSVREHNERQMDKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARRVREVGFKV					
orf75-1		GIQKLVSVREHNERQMDKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARRVREAGFKV					
		70	80	90	100	110	120
orf75a.pep		VPVVGASAVMAALSVAGVAGSDFYFNGFVPPKSGERRKLFKWKVRVAFPVVMFETPHRIG					
orf75-1		VPVVGASAVMAALSVAGVEGSDFYFNGFVPPKSGERRKLFKWKVRAAFPIVMFETPHRIG					
		130	140	150	160	170	180
orf75a.pep		ATLADMAELFPERRMLLAREITKTFETFLSGTVGEIQTALAADGNQSRGEMVVLVLYPAQD					
orf75-1		ATLADMAELFPERRMLLAREITKTFETFLSGTVGEIQTALSADGNQSRGEMVVLVLYPAQD					
		190	200	210	220	230	240
orf75a.pep		EKHEGLSESAQNIMKILTAE LPTKQAAELA AKITGEGKKALYDLALSWKNKX					
orf75-1		EKHEGLSESAQNIMKILTAE LPTKQAAELA AKITGEGKKALYDLALSWKNKX					
		250	260	270	280	290	
orf75a.pep		MSVFQTAFFMFQKHLQKASDSVVGGTLYVVATPIGNLADITLRLALAVLQKADIICAEDTR					
orf75ng		MSVFQTAFFMFQKHLQKASDSVVGGTLYVVATPIGNLADITLRLALAVLQKADIICAEDTR					

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF75 shows 93.2% identity over a 292aa overlap with a predicted ORF (ORF75.ng) from *N. gonorrhoeae*:

60 orf75.pep MFVFQTAFFMFQKHLQKASDSVVGGTLYVVATPIGNLADITLRLALAVLQKADIICAEDTR 56  
 orf75ng MSVFQTAFFMFQKHLQKASDSVVGGTLYVVATPIGNLADITLRLALAVLQKADIICAEDTR 60

5	orf75.pep	VTAQLLSAYGIQGLVSVREHNERQMADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLAR	116
	orf75.ng	VTAQLLSAYGIQGLVSVREHNERQMADKIVGYLSDGLVVAQVSDAGTPAVCDPGAKLAR	120
10	orf75.pep	RVREAGFKVVPVVGASAVMAALSVAGVEGSDFYFNGFVPPKSGERRKLFKQWVRAAFPIV	176
	orf75.ng	RVREAGFKVVPVVGASAVMAALSVAGVAESDFYFNGFVPPKSGERRKLFKQWVRAAFPVV	180
15	orf75.pep	MFETPHRIGALADMAELFPERRMLAREITKTFETFLSGTVGEIQTALSADGQDSRGEM	236
	orf75.ng	MFETPHRIGATLADMAELFPERRMLAREITKTFETFLSGTVGEIQTALAADGNQDSRGEM	240
20	orf75.pep	VLVLYPAQDEKHEGLSESAQNIMKILTAELPTKQAAELAAKITGEGKKALYD	288
	orf75.ng	VLVLYPAQDEKHEGLSESAQNAMKILAAELPTKQAAELAAKITGEGKKALYDLALSWKNK	300

An ORF75ng nucleotide sequence <SEQ ID 291> was predicted to encode a protein having amino acid sequence <SEQ ID 292>:

1	MSVFQTAFFM	FQKHLQKASD	SVVGGTLYV	ATPIGNLADI	TLRALAVLQK
51	ADIICAEDTR	VTAQLLSAYG	IQGLVSVRE	HNERQMADKV	IGFLSDGLV
101	AQVSDAGT	PAVCDPGAKLAR	RVREAGFKV	VPVVGASAVMA	ALSVAGVAES
151	DFYFNGFV	PPKSGERRKLF	KQWVRAAFPV	MFETPHRIG	TLADMAELFP
201	ERRMLAREI	TKTFETFLSG	TVGEIQTALA	ADGNQDSRGEM	VLVLYPAQDE
251	KHEGLSESAQ	NAMKILAAEL	PTKQAAELAA	KITGEGKKAL	YDLALSWKNK
301	*				

After further analysis, the following gonococcal DNA sequence <SEQ ID 293> was identified:

1	ATGTTTCAGA	AACACTTGCA	GAAAGCCTCC	GACAGCGTCG	TCGGAGGGAC
51	ATTATACGTG	GTTGCCACGC	CCATCGGCAA	TTTGGCAGAC	ATTACCTGTC
101	GCGCTTTGGC	GGTATTGCAA	AAGGCGGACA	TCATTTGTGC	CGAAGACACG
151	CGCGTTACTG	CGCAGCTTTT	GAGCGCGTAC	GGCATTTCAGG	GCAGGTTGGT
201	CAGTGTGCGC	GAACACAACG	AGCGGCAGAT	GGCGGACAAG	GTAATCGGTT
251	TCCTTTTCAGA	CGGCCTGGTT	GTGGCGCAGG	TTTCCGATGC	GGGTACGCCG
301	GCCGTGTGCG	ACCCGGGCGC	GAAACTCGCC	CGCCGCGTGC	GCGAAGCAGG
351	GTTCAAAGTC	GTTCCCGTCG	TGGCGCGAAG	CGCGGTAATG	GCGGCGTTGA
401	GTGTGGCCGG	TGTGGCGGAA	TCCGATTTT	ATTTCACCGG	TTTTGTACCG
451	CCGAAATCGG	GCGAACGTAG	GAAATTGTTT	GCCAAATGGG	TGCGGCGCGC
501	ATTTCCTGTC	GTCATGTTTG	AAACGCCGCA	CCGAATCGGG	GCAACGCTTG
551	CCGATATGGC	GGAATTGTTC	CCCGAACGCC	GTCTGATGCT	GGCGCGCGAA
601	ATCACGAAAA	CGTTTGAAAC	GTTCTTAAGC	GGCACGGTTG	GGGAAATTCA
651	GACGGCATTG	GCGGCGGACG	GCAACCAATC	GCGCGGCGAG	ATGGTGTGTTG
701	TGCTTTATCC	GCGCGCAGGAT	GAAAACACG	AAGGCTTGTC	CGAGTCTGCG
751	CAAAATGCGA	TGAAATCCT	TGCGGCGGAG	CTGCCGACCA	AGCAGGCGGC
801	GGAGCTTGCC	GCCAAGATTA	CAGGTGAGGG	CAAAAAGGCT	TTGTACGATT
851	TGGCACTGTC	GTGGAAAAAC	AAATGA		

45 This corresponds to the amino acid sequence <SEQ ID 294; ORF75ng-1>:

1	MFQKHLQKAS	DSVVGGTLYV	VATPIGNLAD	ITLRALAVLQ	KADIICAEDT
51	RVTAQLLSAY	GIQGLVSVR	EHNERQMADK	VIGFLSDGLV	VAQVSDAGTP
101	AVCDPGAKLA	RRVREAGFKV	VPVVGASAVM	AALSVAGVAE	SDFYFNGFVP
151	PKSGERRKLF	AKWVRAAFPV	VMFETPHRIG	ATLADMAELF	PERRMLARE
201	ITKTFETELS	GTVGEIQTAL	AADGNQSRGE	MVLVLYPAQD	EKHEGLSESA
251	QNAMKILAAE	LPTKQAAELA	AKITGEGKKA	LYDLALSWKN	K*

ORF75ng-1 and ORF75-1 show 96.2% identity in 291 aa overlap:

55	orf75-1.pep	10	20	30	40	50	60
	orf75ng-1	10	20	30	40	50	60
60	orf75-1.pep	70	80	90	100	110	120
	orf75ng-1	70	80	90	100	110	120

		:             :::    :
	orf75ng-1	GIQGR LVS VREHNERQMA DKVIGFLSDGLVVAQVSDAGT PAVCDP GAKLARRVREAGFKV
		70 80 90 100 110 120
5		130 140 150 160 170 180
	orf75-1.pep	VPVVGASAVMAALS VAGVEGSDFYFNGFVPPKSGERRKLFAKWVRAAFPVIMFETPHRIG
		:     :     :     :     :     :     :     :     :     :     :     :
	orf75ng-1	VPVVGASAVMAALS VAGVAESDFYFNGFVPPKSGERRKLFAKWVRAAFPVVMFETPHRIG
		130 140 150 160 170 180
10		190 200 210 220 230 240
	orf75-1.pep	ATLADMAELFPERR LMLAREITKTFETFLSGTVGEIQTALSADGNQSRGEMVLVLVPAQD
		:     :     :     :     :     :     :     :     :     :     :
	orf75ng-1	ATLADMAELFPERR LMLAREITKTFETFLSGTVGEIQTALAADGNQSRGEMVLVLVPAQD
		190 200 210 220 230 240
15		250 260 270 280 290
	orf75-1.pep	EKHEGLSESAQNIMKI LTAELPTKQAAELAAKITGEGKKALYDLALSWNKX
		:     :     :     :     :     :     :     :     :     :     :
20	orf75ng-1	EKHEGLSESAQNAMKI LAAELPTKQAAELAAKITGEGKKALYDLALSWNKX
		250 260 270 280 290

Furthermore, ORG75ng-1 shows significant homology to a hypothetical *E.coli* protein:

sp|P45258|YRAL\_ECOLI\_HYPOTHETICAL 31.3 kD PROTEIN IN AGAI-MTR INTERGENIC REGION (F286)  
25 >gi|606086 (U18997) ORF\_f286 [Escherichia coli]  
>gi|1789535 (AE000395) hypothetical 31.3 kD protein in agai-mtr intergenic region [Escherichia coli] Length = 286  
Score = 218 bits (550), Expect = 3e-56  
Identities = 128/284 (45%), Positives = 171/284 (60%), Gaps = 4/284 (1%)  
30  
Query: 4 KHLQKASDSVVGGLTYVATPIGNLADITLRALAVLQKADIICAEDTRVTAQLLSAYGIQ 63  
K Q A + S G QLY+V TPIGNLADIT RAL VLQ D+I AEDTR T LL +GI  
Sbjct: 2 KQHQSADNSQ--GLYIVVPTPIGNLADITQRALEVLQAVDLIAAEDTRHTGLLLQHFGIN 59  
35  
Query: 64 GRLVSVREHNERQMADKVI GFLSDGLVVAQVSDAGT PAVCDPGAKLARRVREAGFKVVPV 123  
RL ++ +HNE+Q A+ ++ L +G +A VSDAGTP + DPG L R REAG +VVP+  
Sbjct: 60 ARLFALHDHNEQQKAETLLAKLQEGQNIALVSDAGTFLINDPGYHLVRTCREAGIRVVPL 119  
40  
Query: 124 VGASAVMAALS VAGVAESDFYFNGFVPPKSGERRKLF AKWVRAAFVVMFETPHRIGATL 183  
G A + ALS AG+ F + GF+P KS RR ++ +E+ HR+ +L  
Sbjct: 120 PGCAAITALS AAGLP SDRFCYEGFLPAKSKGRRDALKAI EAEPRTLIFYESTHRL DLSL 179  
45  
Query: 184 ADMAELFPERR-LMLAREITKT FETFLSGTVGEIQTALAADGNQSRGEMVLVL YPAQDEK 242  
D+ + E R ++LARE+TKT+ET VGE+ + D N+ +GEMVL++ +  
Sbjct: 180 EDIVAVLGESRYVVLARELTKTWETIHGAPVGELLAWVKEDENRRKGEMVLIV-EGHKAQ 238  
Query: 243 HGLSESAQNAMKILAAELPTKQAAELAAKITGEGKKALYDLAL 286  
E L A + +L AELP K+AA LAA+I G K ALY AL  
Sbjct: 239 EEDLPADALRTLALLQAEPLPKKAAALAAEIHGVKKNALYKYAL 282

Based on this analysis, including the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 35

55 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 295>:

```

1  ATGAAACAGA  AAAAAACCGC  TGCCGCAGTT  ATTGCTGCAA  TGTGGCAGG
51  TTTTGCGGCA  GC.AAAGCAC  CCGAAATCGA  CCCGGCTTTG  .....
                               //
651  .....  ...GAGTTGG  TCAGAAACCA  GTGGAGCAG  GGTTTGAGAC

```

701 AGGAAAAAGC CCGCTTGAAA ATCGATGCCC TTTTGAAGA AAACGGTGTC  
751 AAACCGTAA

This corresponds to the amino acid sequence <SEQ ID 296; ORF76>:

1 MKQKKTAAAV IAAMLGFAA XKAPEIDPAL .....  
5  
201 ..... ELVRNQLEQG LRQEKARLKI DALLEENGVK  
251 P\*

Further work revealed the complete nucleotide sequence <SEQ ID 297>:

1 ATGAAACAGA AAAAAACGCG TGCCGCAGTT ATTGCTGCAA TGTTGGCAGG  
10 51 TTTTGCGGCA GCCAAAGCAC CCGAAATCGA CCCGGCTTTG GTGGATACGC  
101 TGGTGGCGCA GATCATGCAG CAGGCAGACC GGCATGCGGA GCAGTCCCAA  
151 AAACCGGACG GGCAGGCAAT CCGAAACGAT GCCGTCCGCC GGCTACAAAC  
201 TTTGGAAGTT TTGAAAAACA GGCATTGAA GGAAGGTTTG GATAAGGATA  
251 AGGATGTCCA AAACCGCTTT AAAATCGCCG AAGCGTCTTT TTATGCCGAG  
15 301 GAGTACGTCC GTTTCTGGA ACGTTCGGAA ACGGTTTCCG AAGACGAGCT  
351 GCACAAGTTT TACGAACAGC AAATCCGCAT GATCAAATTG CAGCAGGTCA  
401 GCTTCGCAAC CGAAGAGGAG GCGCGTCAGG CGCAGCAGCT CCTGCTCAAA  
451 GGGCTGTCTT TTGAAGGGCT GATGAAGCGT TATCCGAACG ACGAGCAGGC  
501 TTTTGACGGT TTCATTATGG CGCAGCAGCT TCCCGAGCCG CTGGCTTCGC  
20 551 AGTTTGCCGC GATGAATCGG GCGACGTTA CCCGCGATCC GGTCAAATTG  
601 GGCGAACGCT ATTATCTGTT CAAACTCAGC GAGGTCGGGA AAAACCCCGA  
651 CGCGCAGCCT TTCGAGTTGG TCAGAAACCA GTTGGAGCAG GGTTCGAGAC  
701 AGGAAAAAGC CCGCTTGAAA ATCGATGCCC TTTTGAAGA AAACGGTGTC  
751 AAACCGTAA

25 This corresponds to the amino acid sequence <SEQ ID 298; ORF76-1>:

1 MKQKKTAAAV IAAMLGFAA AKAPEIDPAL VDTLVAQIMQ QADRHAEQSQ  
51 KPDGQAIRND AVRRLQTLEV LKNRALKEGL DKDKDVQNRK KIAEASFYAE  
101 EYVRFLESE TVSEDELHKF YEQQIRMIKL QQVSFATEEE ARQAQQLLLK  
151 GLSFEGLMKR YPNDEQAFDG FIMAQQLPEP LASQFAAMNR GDVTRDPVKL  
201 GERYYLFLKS EVGKNPDAQP FELVRNQLEQ GLRQEKARLK IDALLEENG  
251 KP\*

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF76 shows 96.7% identity over a 30aa overlap and 96.8% identity over a 31aa overlap with an

35 ORF (ORF76a) from strain A of *N. meningitidis*:

	10	20	30	
orf76.pep	MKQKKTAAAVIAAMLGFAAXKAPEIDPAL			
orf76a	MKQKKTAAAVIAAMLGFAAXKAPEIDPALVDTLVAQIMQQADRHAEQSQKPDGQAIRND			
	10	20	30	40 50 60
	//			
		70	80	90
orf76.pep		XELVRNQLEQGLRQEKARLKIDALLEENGVKPX		
orf76a	DVTRDPVKLG	ERYLFLK	SEVGKNPDAQPFELVRNQLEQGLRQEKARLKID	AILEENGVKPX
	200	210	220	230 240 250

The complete length ORF76a nucleotide sequence <SEQ ID 299> is:

1 ATGAAACAGA AAAAAACGCG TGCCGCAGTT ATTGCTGCAA TGTTGGCAGG  
50 51 TTTTGCGGCA GCCAAAGCAC CCGAAATCGA CCCGGCTTTG GTGGATACGC  
101 TGGTGGCGCA GATCATGCAG CAGGCAGACC GGCATGCGGA GCAGTCCCAA  
151 AAACCGGACG GGCAGGCAAT CCGAAACGAT GCCGTCCGTC GGCTGCAAAAC  
201 TTTGGAAGTT TTGAAAAACA GGCATTGAA GGAAGGTTTG GATAAGGATA  
251 AGGATGTCCA AAACCGCTTT AAAATCGCCG AAGCGTCTTT TTATGCCGAG  
301 GAGTACGTCC GTTTCTGGA ACGTTCGGAA ACGGTTTCCG AAAGCGCACT  
55 351 GCGTCAGTTT TATGAGCGGC AAATCCGCAT GATCAAATTG CAGCAGGTCA

5  
401 GCTTCGCAAC CGAAGAGGAG GCGCGTCAGG CGCAGCAGCT CCTGCTCAAA  
451 GGGCTGTCTT TTGAAGGGCT GATGAAGCGT TATCCGAACG ACGAGCAGGC  
501 TTTTGACGGT TTCATTATGG CGCAGCAGCT TCCCGAGCCG CTGGCTTCGC  
551 AGTTTGCAGC GATGAATCGG GCGCAGCTTA CCCGCGATCC GGTCAAATTG  
601 GCGCAACGCT ATTATCTGTT CAAACTCAGC GAGGTCGGGA AAAACCCCGA  
651 CGCGCAGCCT TTCGAGTTGG TCAGAAACCA GTTGAACAA GGTGTGAGAC  
701 AGGAAAAAGC CCGCTTGAAA ATCGATGCCA TTTTGAAGA AAACGGTGTC  
751 AAACCGTAA

This encodes a protein having amino acid sequence <SEQ ID 300>:

10  
1 MKQKKTAADV IAAMLGFAA AKAPEIDPAL VDTLVAQIMQ QADRHAEQSQ  
51 KPDGQAIRND AVRRLQTLV LKNRALKEGL DKDKDVQNRK KIAEASFYAE  
101 EYVRFLESE TVSESALRQF YERQIRMIKL QQVSFATEEE ARQAQQLLLK  
151 GLSFEGLMKR YPNDEQAFDG FIMAQQLPEP LASQFAAMNR GDVTRDPVKL  
15  
201 GERYLFLKLS EVGKNPDAQF FELVRNQLQGLRQEKARLK IDAILEENG  
251 KP\*

ORF76a and ORF76-1 show 97.6% identity in 252 aa overlap:

20  
10 20 30 40 50 60  
orf76a.pep MKQKKTAADVIAAMLGFAAAXKAPEIDPALVDTLVAQIMQQADRHAEQSQKPDGQAIRND  
|||:|||||  
orf76-1 MKQKKTAADVIAAMLGFAAAXKAPEIDPALVDTLVAQIMQQADRHAEQSQKPDGQAIRND  
10 20 30 40 50 60  
25  
70 80 90 100 110 120  
orf76a.pep AVRRLQTLV LKNRALKEGLDKDKDVQNRKIAEASFYAEYVRFLESE TVSESALRQF  
|||:|||||  
orf76-1 AVRRLQTLV LKNRALKEGLDKDKDVQNRKIAEASFYAEYVRFLESE TVSEDELHKF  
70 80 90 100 110 120  
30  
130 140 150 160 170 180  
orf76a.pep YERQIRMIKLQQVSFATEEEARQAQQLLLKGLSFEGLMKRYPNDEQAFDGFIMAQQLPEP  
||:|||||  
orf76-1 YEQQIRMIKLQQVSFATEEEARQAQQLLLKGLSFEGLMKRYPNDEQAFDGFIMAQQLPEP  
130 140 150 160 170 180  
35  
190 200 210 220 230 240  
orf76a.pep LASQFAAMNRGDVTRDPVKLGERYLFLKSEVGNPDAQPFELVRNQLQGLRQEKARLK  
|||:|||||  
orf76-1 LASQFAAMNRGDVTRDPVKLGERYLFLKSEVGNPDAQPFELVRNQLQGLRQEKARLK  
190 200 210 220 230 240  
40  
250  
orf76a.pep IDAILEENGVKPX  
|||:|||||  
orf76-1 IDALLEENGVKPX  
45  
250

#### Homology with a predicted ORF from *N.gonorrhoeae*

The aligned aa sequences of ORF76 and a predicted ORF (ORF76.ng) from *N. gonorrhoeae* of the N- and C-termini show 96.7 % and 100% identity in 30 and 31 overlap, respectively:

50  
orf76.pep MKQKKTAADVIAAMLGFAAAXKAPEIDPAL 30  
|||:|||||  
orf76ng MKQKKTAADVIAAMLGFAAAXKAPEIDPALVDTLVAQIMQQADRHAEQSQRPDQAIRND 60  
//  
55  
orf76.pep ELVRNQLQGLRQEKARLKIDALLEENGVKP 251  
|||:|||||  
orf76ng VTRNPVKLGERYLFLKLGAVGNPDAQPFELVRNQLQGLRQEKARLKIDALLEENGVKP 251

The complete length ORF76ng nucleotide sequence <SEQ ID 301> is:

60  
1 ATGAAACAGA AAAAGACCGC TGCCGAGTT ATTGCTGCAA TGTTGGCAGG  
51 TTTTGGCGCA GCCAAAGCAC CCGAAATCGA CCCGGCTTTG GTGGATACGC  
101 TGGTGGCGCA GATCATGCAG CAGGCAGACC GGCATGCGGA GCAGTCCCAA

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65  
70  
75

```

151 AGACCGGACG GGCAGGCAAT CCGAAACGAT GCCGTCCGCC GGCTGCAAAC
201 TTTGGAAGTT TTGAAAAACA GGGCATTGAA GGAAGGTTTG GATAAGGATA
251 AGGATGTCCA AAACCGCTTT AAAATCGCCG AAGCGTCTTT TTATGCCGAG
301 GAGTACGTCC GTTTTCTGGA ACGTTCGGAA ACGGTTCCG AAAGCGCACT
351 GCGTCAGTTT TATGAGCGGC AAATCCGCAT GATCAAATTG CAGCAGGTCA
401 GCTTCGCAAC CGAAGAGGAG GCGCGTCAGG CGCAGCAGCT CCTGCTCAAA
451 GGGCTGTCTT TTGAAGGGCT GATGAAGCGT TATCCGAACG ACGAGCAGGC
501 GTTCGACGGT TTCATTATGG CGCAGCAGCT TCCCGAGCCG CTGGCTTcgc
551 agtttgCCGG TATGAACCGT GGCGACGTTA CCCGCAATCC GGTCAAATTG
601 GCGAACGCT ATTACCTGTT CAAACTCGGC GCGGTCGGGA AAAACCCCGA
651 CGCGCAGCCT TTCGAGTTGG TCAGAAACCA GTTGAACAA GGTTCGAGGC
701 AGGAAAAAGC CCGCTTGAAA ATCGATGCCC TTTTGGAAga Aaacggtgtc
751 AaacCGTAA

```

This encodes a protein having amino acid sequence <SEQ ID 302>:

15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65  
70  
75

```

1 MKQKKTAAAV IAAMLAGFAA AKAPEIDPAL VDTLVAQIMQ QADRHAEQSQ
51 RPDGQAIRND AVRRLQTLEV LKNRALKEGL DKDKDVQNRF KIAEASFYAE
101 EYVRFLERSE TVSESALRQF YERQIRMIKL QQVSFATEEE ARQAQQLLLK
151 GLSFEGLMKR YPNDEQAFDG FIMAQQLPEP LASQFAGMNR GDVTRNPVKL
201 GERYYLFLKG AVGKNPDAQF FELVRNQLEQ GLRQEKARLK IDALLEENGV
251 KP*

```

ORF76ng and ORF76-1 show 96.0% identity in 252 aa overlap

25  
30  
35  
40  
45  
50  
55  
60  
65  
70  
75  
80  
85  
90  
95  
100  
105  
110  
115  
120  
125  
130  
135  
140  
145  
150  
155  
160  
165  
170  
175  
180  
185  
190  
195  
200  
205  
210  
215  
220  
225  
230  
235  
240  
245  
250

```

orf76-1.pep MKQKKTAAAVIAAMLAGFAAAKAPEIDPALVDTLVAQIMQQADRHAEQSQKPDGQAIRND
orf76ng MKQKKTAAAVIAAMLAGFAAAKAPEIDPALVDTLVAQIMQQADRHAEQSQRPDGQAIRND

orf76-1.pep AVRRLQTLEVLKNRALKEGLDKDKDVQNRFKIAEASFYAEEYVRFLERSETVSEDELHKF
orf76ng AVRRLQTLEVLKNRALKEGLDKDKDVQNRFKIAEASFYAEEYVRFLERSETVSESALRQF

orf76-1.pep YEQQIRMIKLQQVSFATEEEARQAQQLLLKGLSFEGLMKRYPNDEQAFDGFIMAQQLPEP
orf76ng YERQIRMIKLQQVSFATEEEARQAQQLLLKGLSFEGLMKRYPNDEQAFDGFIMAQQLPEP

orf76-1.pep LASQFAAMNRGDVTRDPVKLGERYLFLKSEVGKNPDAQPFELVRNQLEQGLRQEKARLK
orf76ng LASQFAGMNRGDVTRNPVKLGERYLFLKGAVGKNPDAQPFELVRNQLEQGLRQEKARLK

orf76-1.pep IDALLEENGVKPKX
orf76ng IDALLEENGVKPKX

```

Furthermore, ORF76ng shows significant homology to a *B.subtilis* export protein precursor:

55  
60

```

sp|P24327|PRSA BACSU PROTEIN EXPORT PROTEIN PRSA PRECURSOR >gi|98227|pir||S15269
33K lipoprotein - Bacillus subtilis >gi|39782 (X57271) 33kDa lipoprotein
[Bacillus subtilis]
>gi|2226124|gnl|PID|e325181 (Y14077) 33kDa lipoprotein [Bacillus subtilis]
>gi|2633331|gnl|PID|e1182997 (Z99109) molecular chaperonin [Bacillus subtilis]
Length = 292
Score = 50.4 bits (118), Expect = 1e-05
Identities = 48/199 (24%), Positives = 82/199 (41%), Gaps = 32/199 (16%)

Query: 70 VLKNRALKEGLDK-----DKDVQNRFKIAEASF-----YAEEYVRFLERSETVSE 114
VL ++ LDK DK++ N+ K + Y ++Y++ + E +++
Sbjct: 53 VLTQLVQEKVLDKKYKVSDKEIDNKLKEYKTQLGDQYTALEKQYGKDYLKEQVKYELLTQ 112

```



Query: 115 SA-----LRQFYERQIRMIKLOQVSFATEEEARQAQQLLKGLSFEGLMKRYPN 163  
           A          +++++E      I+   +  A ++ A + ++ L KG FE L K Y  
 Sbjct: 113 KAAKDNKIVTDADIKEYWEGLKGIKIRASHILVADKKTAEEVEKKLKKGEKFEDLAKEYST 172  
 Query: 164 DEQAFDG-----FIMAQQLPEPLASQFAAMNRGDVTRDPVKLGERYYLFLKSEVGKNPDA 218  
           D  A  G      F      Q+ E +      +  G+V+ DPVK     Y++ K +E     D  
 Sbjct: 173 DSSASKGGDLGWFAKEGQMDETFSKAAFKLKTGEVS-DPVKTQYGYHIIKKTEERGKYDD 231  
 Query: 219 QPFELVRNQLEQGLRQEKA 237  
           EL      LEQ L     A  
 Sbjct: 232 MKKELKSEVLEQKLNDNAA 250

Based on this analysis, including the presence of a putative leader sequence and a RGD motif in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF76-1 (27.8kDa) was cloned in the pET vector and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 10A shows the results of affinity purification of the His-fusion protein, Purified His-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 10B), ELISA (positive result), and FACS analysis (Figure 10C). These experiments confirm that ORF76-1 is a surface-exposed protein, and that it is a useful immunogen.

### Example 36

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 303>:

1 ATGAAAAAAT CTTTCCTTAC GCTTGTTCTG TATTCGTCTT TACTTACCGC  
 51 CAGCGAAATT GCCTTACCCC TTGGAATTGG GGATTGAAAC CTTACCGGCG  
 101 GCAAAAATTG CGGAAACGTT TGCCTGACA TTTGTGATG CTGCGCTGTA  
 151 TCTGTTTGCG CGTAATAAGG TGACGCGTTT GTTGATTGCG GTGTTTTTTG  
 201 CGTTCAGCAT TATTGCCAAC AATGTGCATT ACGCGGATTA TCAAAGCTGG  
 251 ATGACG....  
 //  
 1201 ..... CAAACCGTAT TCGAGCAGCT GCAAAGACT CCTGACGGCA  
 1251 ACTGGCTGTT TGCCATATAC TCCGATCATG GCCAGTATGT TCGCCAAGAT  
 1301 ATCTACAATC AAGGCACGGT GCAGCCCGAC AGCTATCTCG TGCCGCTAGT  
 1351 GTTGACAGC CCGGATAAGG CCGTGCAACA GGCTGCCAAC CAGGCTTTTG  
 1401 CGCCTTGCGA GATTGCCTTC CATCAGCAGC TTTCACGTT CCTGATTCAC  
 1451 ACGTTGGGCT ACGATATGCC GGTTCAGGT TGTGCGAAG GCTCGGTAAC  
 1501 GGGCAACCTG ATTACGGGTG ATGCAGGCAG CTTGAACATT CGCGACGGCA  
 1551 AGGCGGAATA TGTTTATCCG CAATGA

This corresponds to the amino acid sequence <SEQ ID 304; ORF81>:

1 MKKSFLTLVL YSSLLTASEI AYPLELGIET LPAKIAETF ALTFVIAALY  
 51 LFARNKVTRL LIAVFFAFSI IANNVHYADY QSWMT.....  
 //  
 401 ...QTVFEQL QKTPDGNWLF AYTSDHGQYV RQDIYNQGTV QPDSYLVPLV  
 451 LYSPDKAVQQ AANQAFAPCE IAFHQQLSTF LIHTLGVDMP VSGCREGSVT  
 501 GNLTGDAGS LNIRDGKAEY VYPQ\*

Further work revealed the complete nucleotide sequence <SEQ ID 305>:

1 ATGAAAAAAT CTTTCCTTAC GCTTGTTCTG TATTCGTCTT TACTTACCGC  
 51 CAGCGAAATT GCCTATCGCT TTGTATTTGG GATTGAAACC TTACCGGCGG

5  
10  
15  
20  
25  
30

```

101 CAAAAATTGC GGAAACGTTT GCGCTGACAT TTGTGATTGC TGCCTGTAT
151 CTGTTTTCGC GTTATAAGGT GACGCGTTTG TTGATTGCGG TGTTTTTTGC
201 GTTCAGCATT ATTGCCAACA ATGTGCATTA CGCGGTTTAT CAAAGCTGGA
251 TGACGGGCAT CAATTATTGG CTGATGCTGA AAGAGGTTAC CGAAGTCGGC
301 AGCGCGGGTG CGTCGATGTT GGATAAGTTG TGGCTGCCTG TGTGTGGGG
351 CGTGTTGGAA GTCATGTTGT TTTGCAGCCT TGCCAAGTTC CGCCGTAAGA
401 CGCATTTTTC TGCCGATATA CTGTTTGCTT TCCTAATGCT GATGATTTTC
451 GTGCGTTTCG TCGACACGAA ACAAGAGCAC GGTATTTTCG CCAAACCGAC
501 ATACAGCCGC ATCAAAGCCA ATTATTTTCT CTTGCGTTAT TTTGTCGGAC
551 GCGTGTTGCC GTATCAGTTG TTTGATTTAA GCAGGATTC CGCCTTTAAG
601 CAGCCTGCTC CAAGCAAAAT CGGCGAGGCG AGTGTTCAAA ATATCGTCCT
651 GATTATGGGC GAAAGCGAAA GCGCGGCGCA TTTGAAGCTG TTTGGCTACG
701 GACGCGAAAC TTCGCCGTTT TTAACCCGCG TGTCGCAAGC CGATTTTAAAG
751 CCGATTGTGA AACAAAGTTA TTCCGAGGCG TTTATGACTG CAGTGTCCCT
801 GCCCAGTTT TCAATGCGA TACGCGACGC CAACGGCTTG GAACAAATCA
851 GCGCGGCGCA TACCAATATG TTCCGCTCG CCAAAGAGCA GGGCTATGAA
901 ACGTATTTT ACAGCGCGCA GCGGAAAAC GAGATGGCGA TTTTGAAGCT
951 AATCGGTAAG AAATGGATAG ACCATCTGAT TCAGCCGACG CAACTTGGCT
1001 ACGGCAACGG CGACAATATG CCCGATGAGA AGCTGCTGCC GTGTTCGAC
1051 AAAATCAATT TGCAGCAGGG CAAGCATTTT ATCGTGTGTC ACCAACGCGG
1101 TTCGCACGCC CCATACGGCG CATGTGTGCA GCCTCAAGAT AAAGTATTCG
1151 GCGAAGCCGA TATTGTGGAT AAGTACGACA ACACCATCCA CAAAACCGAC
1201 CAAATGATTC AAACCGTATT CGAGCAGCTG CAAAAGCAGC CTGACGGCAA
1251 CTGGCTGTTT GCCTATACCT CCGATCATGG CAGTATGTT CGCCAAGATA
1301 TCTACAATCA AGGCACGGTG CAGCCCGACA GCTATCTCGT GCCGCTAGTG
1351 TTGTACAGCC CGGATAAGGC CGTGCAACAG GCTGCCAACC AGGCTTTTTC
1401 GCCTTGCGAG ATTGCCTTCC ATCAGCAGCT TTCAACGTTT CTGATTACAC
1451 CGTTGGGCTA CGATATGCCG GTTTCAGGTT GTCGCGAAGG CTCGGTAACG
1501 GGCAACCTGA TTACGGGTGA TGCAGGCAGC TTGAACATTC GCGACGGCAA
1551 GCGGAATAT GTTTATCCGC AATGA

```

This corresponds to the amino acid sequence <SEQ ID 306; ORF81-1>:

35  
40

```

1 MKKSFLTLVL YSSLLTASEI AYREVFGEIET LPAAKIAETF ALTFVIAALY
51 LFARYKVTRL LIAVFFAFSI IANNVHYAVY QSWMTGINYW LMLKEVTEVG
101 SAGASMLDKL WLPVLWGVLE VMLFCSLAKF RRKTHFSADI LFAFLMLMIF
151 VRSFDTKQEH GISPKPTYSR IKANYFSFGY FVGRVLPYQL FDLSRIPAFK
201 QPAPSKIGQG SVQNIIVLIMG ESESAAHLKL FGYGRETSPP LTRLSQADFK
251 PIVKQSYSAG FMTAVSLPSF FNAIPHANGL EQISGGDTNM FRLAKEQGYE
301 TYFYSAQAEI EMAILNLIGK KWIDHLIQT QLGYGNGDNM PDEKLLPLFD
351 KINLQQGKHF IVLHQRGSHA PYGALLQPD KVFGEADIVD KYDNTIHKTD
401 QMIQTVFEQL QKQPDGNWLF AYTSDHGQYV RQDIYNQGTQ QPDSYLVPLV
451 LYSPPKAVQQ AANQAFAPCE IAFHQQLSTF LIHTLGVDMP VSGCREGSVT
501 GNLTITGDAGS LNIRDGKAEY VYPQ*

```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

45 ORF81 shows 84.7% identity over a 85aa overlap and 99.2% identity over a 121aa overlap with an ORF (ORF81a) from strain A of *N. meningitidis*:

50

```

      10      20      30      40      50      60
orf81.pep  MKKSFLTLVL YSSLLTASEI AYREVFGEIET LPAAKIAETF ALTFVIAALY LFARNKVTRL
          |||:::| |||||:||||| : :|||:|||||:|||||:||||| :|||
orf81a     MKKSFLVLFL YSSLLTASEI AYREVFGEIET LPAAKMAETF ALTFVIAALY LFARYKATRL
          10      20      30      40      50      60

      70      80
orf81.pep  LIAVFFAFSI IANNVHYADY QSWMT
          |||:|||||:||||| |||:|
orf81a     LIAVFFAFSI IANNVHYAVY QSWITGINYWL MLKEITEVGG AGASMLDKLWLPALWGVLE
          70      80      90      100     110     120
                //

      120     130     140
orf81.pep  QTVFEQLQKTPDGNWLFAYTSDHGQYVRQD
          |||:|||||:|||||:|||||:|||||
orf81a     IPHANGLEQISGGDIVDKYDNTIHKTDQMIQTVFEQLQKQPDGNWLFAYTSDHGQYVRQD

```

-221-

		280	290	300	310	320	330
		150	160	170	180	190	200
5	orf81.pep	IYNQGTVPDSYLVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTFLIHTLGYDMPVSG					
	orf81a	IYNQGTVPDSYLVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTFLIHTLGYDMPVSG					
		340	350	360	370	380	390
10	orf81.pep	210	220	230			
	orf81a	CREGSVTGNLITGDAGSLNIRDGKAEYVYPQX					
		400	410	420			

The complete length ORF81a nucleotide sequence <SEQ ID 307> is:

15	1	ATGAAAAAAT	CCCTTTTCGT	TCTCTTTCTG	TATTCGTCCC	TACTTACTGC
	51	CAGCGAAATT	GCTTATCGCT	TTGTATTCCG	AATTGAAACC	TTACCGGCTG
	101	CAAAAATGGC	AGAAACGTTT	GCGCTGACAT	TTGTGATTGC	TGCGCTGTAT
	151	CTGTTTGCGC	GTTATAAGGC	AACGCGTTTG	TTGATTGCGG	TGTTTTTCGC
	201	GTTCAGCATT	ATTGCCAACA	ATGTGCATTA	CGCGGTTTAT	CAAAGCTGGA
20	251	TAACGGGCAT	TAATTATTGG	CTGATGCTGA	AAGAGATTAC	CGAAGTTGGC
	301	GGCGCAGGGG	CGTCGATGTT	GGATAAGTTG	TGGCTGCCTG	CGTTGTGGGG
	351	CGTGTGTGAA	GTTCATGTTGT	TTTGCAGCCT	TGCCAAGTTC	CGCCGTAAGA
	401	CGCATTTTTT	TGCCGATATA	CTGTTTGCTT	TCCTAATGCT	GATGATTTTC
25	451	GTGCGTTCGT	TCGACACGAA	ACAAGAACAC	GGTATTTTCG	CCAAACCGAC
	501	ATACAGCCGC	ATCAAAGCCA	ATTATTTTCA	CTTCGGTTAT	TTTGTCCGAC
	551	GCGTGTGCGC	GTATCAGTTG	TTTGATTTAA	GCAAGATTCC	TGTGTTCAAA
	601	CAGCCTGCTC	CAAGCAGAAT	CGGGCAAGGC	AGTATTCAAA	ATATCGTCCT
	651	GATTATGGGC	GAAAGCGAAA	GCGCGGCGCA	TTTGAAATG	TTTGCTACG
	701	GGCGCGAAAC	TTCCGCGTTT	TTGACCCAGC	TTTCGCAAGC	CGATTTTAAG
30	751	CCGATTGTGA	AACAAAGTTA	TTCCGCGAGC	TTTATGACGG	CAGTATCCCT
	801	GCCCAGTTTC	TTTAACGTCA	TACCGCATGC	CAACGGCTTG	GAACAAATCA
	851	GCGCGGCGCA	TATTGTGGAT	AAGTACGACA	ACACCATCCA	CAAAACCGAC
	901	CAAATGATTC	AAACCGTATT	CGAGCAGCTG	CAAAGCAGC	CTGACGGCAA
	951	CTGGCTGTTT	GCCTATACCT	CCGATCATGG	CCAGTATGTT	CGCCAAGATA
35	1001	TCTACAATCA	AGGCACGGTG	CAGCCCGACA	GCTATCTCGT	GCCGCTGGTG
	1051	TTGTACAGCC	CGGATAAGGC	CGTGCAACAG	GCTGCCAACC	AGGCTTTTGC
	1101	GCCTTGCGAG	ATTGCCTTCC	ATCAGCAGCT	TTCAACGTTT	CTGATTACAC
	1151	CGTTGGGCTA	CGATATGCCG	GTTTCAGGTT	GTCGCGAAGG	CTCGGTAACG
40	1201	GGCAACCTGA	TTACGGGTGA	TGCAGGCAGC	TTGAACATTC	GCGACGGCAA
	1251	GGCGGAATAT	GTTTATCCGC	AATGA		

This encodes a protein having amino acid sequence <SEQ ID 308>:

	1	MKKSFLVFL	YSSLLTASEI	AYRFVFGIET	LPAAKMAETF	ALTFVIAALY
	51	LFARYKATRL	LIAVFFAFSI	IANNVHYAVY	QSWITGINYW	LMLKEITEVG
45	101	GAGASMLDKL	WLPALWGVLE	VMLFCSLAKF	RRKTHFSADI	LFAFLMLMIF
	151	VRSFDTKQEH	GISPKPTYSR	IKANYFSFGY	FVGRVLPYQL	FDLSKIPVFK
	201	QPAPSRIGQG	SIQNIIVLIM	ESESAAHLKL	FGYGRETSPF	LTQLSQADFK
	251	PIVKQSYSAG	FMTAVSLPSF	FNVIPHANG	EQISGGDIVD	KYDNTIHKTD
	301	QMIQTVFEQL	QKQPDGNWLF	AYTSDHGQYV	RQDIYNQGTV	QPDYLVPLV
50	351	LYSPDKAVQQ	AANQAFAPCE	IAFHQQLSTF	LIHTLGVDMP	VSGCREGSVT
	401	GNLITGDAGS	LNIRDGKAEY	VYPQ*		

ORF81a and ORF81-1 show 77.9% identity in 524 aa overlap:

		10	20	30	40	50	60
	orf81a.pep	MKKSFLVFLYSSLLTASEIAYRFVFGIETLPAAKMAETFALTFVIAALYLFARYKATRL					
55	orf81-1	MKKSFLTLVLYSSLLTASEIAYRFVFGIETLPAAKIAETFALTFVIAALYLFARYKVTRL					
		10	20	30	40	50	60
60	orf81a.pep	LIAVFFAFSIIANNVHYAVYQSWITGINYWLMLKEITEVGGAGASMLDKLWLPALWGVLE					
	orf81-1	LIAVFFAFSIIANNVHYAVYQSWMTGINYWLMLKEVTEVGSAGASMLDKLWLPVWGVLE					
		70	80	90	100	110	120

		130	140	150	160	170	180
	orf81a.pep	VMLFCSLAKFRKTHFSADILFAFLMLMIFVRSFDTKQEHGISPKPTYSRKANYFSFGY					
	orf81-1	VMLFCSLAKFRKTHFSADILFAFLMLMIFVRSFDTKQEHGISPKPTYSRKANYFSFGY					
5		130	140	150	160	170	180
	orf81a.pep	FVGRVLPYQLFDLSKIPVFKQPAPSRIGQGSIGNIVLIMGESESAHLKLFYGYGRETSPF					
10	orf81-1	FVGRVLPYQLFDLSRIPAFKQPAPSKIGQGSVQNIIVLIMGESESAHLKLFYGYGRETSPF					
		190	200	210	220	230	240
	orf81a.pep	LTQLSQADFKPIVKQYSAGFMTAVSLPSFFNVIPHANGLEQISGGD-----					
15	orf81-1	LTRLNQADFKPIVKQYSAGFMTAVSLPSFFNAIPHANGLEQISGGDTNMFRLAKEQGYE					
		250	260	270	280	290	300
20	orf81a.pep	-----					
	orf81-1	TYFYSAQAENEMAILNLIGKKWIDHLIQPTQLGYGNGDNMPDEKLLPLFDKINLQQGKHF					
		310	320	330	340	350	360
25	orf81a.pep	-----		290	300	310	320
	orf81-1	IVLHQRGSHAPYGALLQPDQKVFGEADIVDKYDNTIHKTDQMIQTVFEQLQKQPDGNWLF					
30		370	380	390	400	410	420
	orf81a.pep	AYTSDHGCQYVRQDIYNQGTVPDSYLVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTF					
35	orf81-1	AYTSDHGCQYVRQDIYNQGTVPDSYLVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTF					
		430	440	450	460	470	480
	orf81a.pep	LIHTLGYDMPVSGCREGSVTGNLITGDAGSLNIRDGKAEYVYPQX					
40	orf81-1	LIHTLGYDMPVSGCREGSVTGNLITGDAGSLNIRDGKAEYVYPQX					
		490	500	510	520		

### Homology with a predicted ORF from *N.gonorrhoeae*

The aligned aa sequences of ORF81 and a predicted ORF (ORF81.ng) from *N. gonorrhoeae* of the

45 N- and C-termini show 82.4 % and 97.5% identity in 85 and 121 overlap, respectively:

	orf81.pep	MKKSFLTIVLYSSLLTASEIAYPLELGIETLPAAKIAETFALT FVIAALYLFARNKVTRL	60
	orf81ng	MKKSFLVFLYSSLLTASEIAYRFVFGIETLPAAKMAETFALT FVIAALYLFARYKASRL	60
50	orf81.pep	LIAVFFAFSIIANNVHYADYQSWMT	85
	orf81ng	LIAVFFAFSMIANNVHYAVYQSWMTGINYWLMLKEVTEVGSAGASMLDKLWLPALWGVAE	120
		//	
55	orf81.pep	QTVFEQLQKTPDGNWLFAYTSDHGCQYVRQD	433
	orf81ng	ALLQPQDKVFGEADIVDKYDNTIHKTDQMIQTVFEQLQKQPDGNWLFAYTSDHGCQYVRQD	433
	orf81.pep	IYNQGTVPDSYLVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTFLIHTLGYDMPVSG	493
60	orf81ng	IYNQGTVPDSYIVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTFLIHTLGYDMPVSG	493
	orf81.pep	CREGSVTGNLITGDAGSLNIRDGKAEYVYPQ	524
	orf81ng	CREGSVTGNLITGDAGSLNIRNGKAEYVYPQ	524

65 The complete length ORF81ng nucleotide sequence <SEQ ID 309> is:

[illegible]

35

40

50

55

60

65

	orf81-1	FVGRVLPYQLFDLSRIPAFKQPAPSKIGQGSVQNIIVLIMGESESAHLKLFYGRGRETSPF
		190 200 210 220 230 240
5	orf81ng-1.pep	LTRLAQADFKPIVKQSYSAGFMTAVSLPSFFNVIPHANGLEQISGGDTNMFRLAKEQGYE
	orf81-1	LTRLAQADFKPIVKQSYSAGFMTAVSLPSFFNAIPHANGLEQISGGDTNMFRLAKEQGYE
		250 260 270 280 290 300
10	orf81ng-1.pep	TYFYSAQAENQMAILNLIGKKWIDHLIQPTQLGYGNGDNMPDEKLLPLFDKINLQQGRHF
	orf81-1	TYFYSAQAENQMAILNLIGKKWIDHLIQPTQLGYGNGDNMPDEKLLPLFDKINLQQGRHF
		310 320 330 340 350 360
15	orf81ng-1.pep	IVLHQRGSHAPYGALLQPQDKVFGADIVDKYDNTIHKTDQMIQTVEFQQLQKQPDGNWLF
	orf81-1	IVLHQRGSHAPYGALLQPQDKVFGADIVDKYDNTIHKTDQMIQTVEFQQLQKQPDGNWLF
		370 380 390 400 410 420
20	orf81ng-1.pep	AYTSDHGGQYVRQDIYNQGTVPDSYIVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTF
	orf81-1	AYTSDHGGQYVRQDIYNQGTVPDSYIVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTF
		430 440 450 460 470 480
25	orf81ng-1.pep	LIHTLGYDMPVSGCREGSVTGNLITGDAGSLNIRNGKAIEYVYPQX
	orf81-1	LIHTLGYDMPVSGCREGSVTGNLITGDAGSLNIRNGKAIEYVYPQX
		490 500 510 520

Furthermore, ORF81ng shows significant homology to an *E.coli* OMP:

35	gi 1256380 (U50906) outer membrane adherence protein-associated protein [E. coli] Length = 547 Score = 87.4 bits (213), Expect = 2e-16 Identities = 122/468 (26%), Positives = 198/468 (42%), Gaps = 70/468 (14%)
40	Query: 25 VFGIETLPAKMAETFA-LTFMIAALYLFARYKAS--RLLIIVFFAFSMIANNVHYAVYQ 81 VFGI L A+ A L F + + + R + RLL+A F + A ++ ++Y Sbjct: 29 VFGITNLVASSGAHMQRLFFVLTLIVVKRISSPLRLVAAPFVL-LTAADMSISLY- 86
45	Query: 82 SWMT-----GINYWLMLKEVTEVGSAGASMLDKLWLPALWGVAEVMFLFCSLAKFRKRT 134 SW T G ++ + EV A ML ++ P L A + L + Sbjct: 87 SWCTFGTTFNDGFAISVLQSDPDEV---AKMLG-MYSPYLCAFAFLSLLFLAVIICYDV 141
50	Query: 135 HFSADILFAFLMLMIFVRSF-----DTKQEHGISPCKPTYSRIKAN--YFSFGYFVG 183 + L+L++ S D K ++ SP SR +F+ YF Sbjct: 142 SLPTKKVTGILLIVISGSLFSACQFAYKDAKNKNAFSPYILASRFATYTPFFNLNYFAL 201
55	Query: 184 RVLPHYQ--LFDLSKIPVFKQPAPSKIGQGSIQNIIVLIMGESESAHLKLFYGRGRETSPFL 241 +Q L + +P F+ + I VLI+GES ++ L+GY R T+P + Sbjct: 202 AAKEHQRLLSIANTVPYFQL----SVRDTGIDTYVLIVGESVRVDNMSLYGYTRSTTPQV 257
60	Query: 242 TRLSQADFKPIVKQSYSAGFMTAVSLP---SFFNVIPHANGLEQISGGDTNMFRLAKEQG 298 +Q + Q+ S TA+S+P + +V+ H I N+ +A + G Sbjct: 258 E--AQRKQIKLFNQAISGAPYTALSVPLSLTADSVLSH-----DIHNYPDNIINMANQAG 310
65	Query: 299 YETFYSAQA---ENQMAILNLIGKKWIDHLIQPTQLGYGNGDNMPDEKLLPLFDKINLQ 355 ++T++ S+Q+ +N A+ ++ ++ + Y G DE LLP + Q Sbjct: 311 FQTFWLSSQSAFRQNGTAVTSI-----AMRAMETVYVRGF---DELLPHLSQALQQ 359
	Query: 356 --QGRHFIVLHQRGSHAPYGALLQPQDKVFGADIVDK-YDNTIHKTDQMIQTVEFQQLQ 412 Q + IVLH GSH P + VF D D YDN+IH TD ++ VFE L+ Sbjct: 360 NTQQKKLIVLHLNGSHEPACSAYPQSSAVFQPDQDQACYNDSIHYTDSLGLQGVFELLK- 418
	Query: 413 QPDGNWLFAYTSDHG---QYVRQDIYNQGT--TVQPDYIVPL-VLYSP 454 D Y +DHG +++++Y G +Y VP+ + YSP

Sbjct: 419 --DRRASVMYFADHGLERDPTKKNVYFHGGREASQQAYHVPFIWYSP 464

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 37

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 311>:

```

10      1  ...ACCCTGCTCC TCTTCATCCC CCTCGTCCTC ACAC.GTGGC GCACACTGAC
      51  CGGCATACTC GCCCACGGCG GCGGCAAACG CTTTGCCGTC GAACAAGAAC
     101  TCGTCGCCGC ATCGTCCCGC GCCGCCGTCA AAGAAATGGA TTTGTCCGCC
     151  yTAAAGGAC GCAAAGCCCG CyTTTACGTC TCCGTATGCG GCGACCAAGG
     201  TTCGGGCAAC ATAAGCGGCG GACGCTACTC TATCGACGCA CTGATACGCG
     251  GCGGCTACCA CAACAACCCC GAAAGTGCCA CCAATACAG CTACCCCGCC
     301  TACGACACTA CCGCCACCAC CAAATCCGAC GCGCTCTCCA GCGTAACCAC
     351  TTCCACATCG CTTTTGAACG CCCC GCCCGC CGyCyTGACG AAAACACGCG
     401  GACGCAAAGG CGAACGcTCC GCCGACTGT CCGTCAACGG CACGGGCGAC
     451  TACCGCAACG AAACCCTGCT CGCCAACCCC CGCGACGTTT CCTTCCTGAC
     501  CAACCTCATC CAAACCGTCT TCTACCTGCG CGGCATCGAA GTCgTACCGC
     551  CCGrATACGC CGACACCGAC GTATTCGTAA CCGTCGACGT A...
```

This corresponds to the amino acid sequence <SEQ ID 312; ORF83>:

```

25      1  ..TLLFIPLVL TXCGTLTGIL AHGGGKRFAV EQELVAASSR AAVKEMDLA
     51  LKGRKAAXYV SVMGDQSGN ISGGRYSIDA LIRGGYHNNP ESATQYSYPA
     101  YDTTATTKSD ALSSVTSTLS LLNAPAAXLT KNSGRKGRS AGLSVNGTGD
     151  YRNETLLANP RDVSFLTNLI QTVFYLRGIE VVPPXYADTD VFVTVDV..
```

Further work revealed the complete nucleotide sequence <SEQ ID 313>:

```

30      1  ATGAAACCC TGCTCCTCCT CATCCCCCTC GTCCTCACAG CCTGCGGCAC
     51  ACTGACCGGC ATACCCGCCC ACGGCGGCGG CAAACGCTTT GCCGTGCAAC
     101  AAGAACTCGT CGCCGCATCG TCCGCGCGCG CCGTCAAAGA AATGGATTG
     151  TCCGCCCTAA AAGGACGCAA AGCCGCCCTT TACGTCTCCG TTATGGGCGA
     201  CCAAGGTTTCG GGCAACATAA GCGGCGGACG CTA CTCTATC GACGCACTGA
     251  TACGCGGCGG CTACCACAAC AACCCCGAAA GTGCCACCCA ATACAGCTAC
     301  CCCGCCCTACG AACTACCGC CACCACCAAA TCCGACGCGC TCTCCAGCGT
     351  AACCCTTCC ACATCGCTTT TGAACGCCCC CGCCGCGGCC CTGACGAAAA
     401  ACAGCGGACG CAAAGGCGAA CGCTCCGCGG GACTGTCCGT CAACGGCACG
     451  GGC GACTACC GCAACGAAAC CCTGCTCGCC AACCCCGCGG ACGTTTCCTT
     501  CCTGACCAAC CTCATCCAAA CCGTCTTCTA CCTGCGCGGC ATCGAAGTCG
     551  TACCGCCCGA ATACGCGGAC ACCGACGTAT TCGTAACCGT CGACGTATTC
     601  GGCACCGTCC GCAGCCGTAC CGAACTGCAC CTCTACAACG CCGAAACCTT
     651  TAAAGCCCAA ACCAAGCTCG AATATTTTCG CTTTGACCGC GACAGCCGGA
     701  AACTGCTGAT TACCCCTAAA ACCGCGGCGT ACGAATCCCA ATACCAAGAA
     751  CAATACGCCC TTTGGACCGG CCCTTACAAA GTCAGCAAAA CCGTCAAAGC
     801  CTCAGACCGC CTGATGGTCG ATTCTCCGA CATTACCCCG TACGGCGACA
     851  CAACCGCCCA AAACCGTCCC GACTTCAAAC AAAACAACGG TAAAAAACCC
     901  GATGTCGGCA ACGAAGTCAT CCGCGCGCGC AAAGGAGGAT AA
```

This corresponds to the amino acid sequence <SEQ ID 314; ORF83-1>:

```

50      1  MKTLLLLIPL VLTACGTLTG IPAHGGGKRF AVEQELVAAS SRAAVKEMDL
     51  SALKGRKAAL YVSVMGDQGS GNISGGRYSI DALIRGGYHN NPESATQYSY
     101  PAYDTTATTK SDALSSVTTS TSLNAPAAA LTKNSGRKGE RSAGLSVNGT
     151  GDYRNETLLA NPRDVSFLTN LIQTVFYLRG IEVVPPEYAD TDVFVTVDFV
     201  GTVRSRTELH LYNAETLKAQ TKLEYFAVDR DSRKLLITPK TAAYESQYQE
     251  QYALWTGPYK VSKTVKASDR LMFVDFSDITP YGDTTAQNRP DFKQNNKKPK
     301  DVGNEVIRRR KGG*
```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF83 shows 96.4% identity over a 197aa overlap with an ORF (ORF83a) from strain A of *N. meningitidis*:

5		10	20	30	40	50
	orf83.pep	TLLFIPLVLT	XCGTLTGILAHGGGKRFAVEQELVAASSRAAVKEMDLSALKGRKAAX			
	orf83a	MKTLXLPLVLT	ACGTLTGIPAHGGGKRFAVEQELVAASSRAAVKEMDLSALKGRKAAL			
10		60	70	80	90	100
	orf83.pep	YVSVMGDQGS	GNISGGGRYSIDALIRGGYHNNPESATQYSYPAYDTTATTKSDALSSVTTS			
	orf83a	YVSVMGDQGS	GNISGGGRYSIDALIRGGYHNNPESATQYSYPAYDTTATTKSDALSSVTTS			
15		70	80	90	100	110
	orf83.pep	120	130	140	150	160
	orf83a	120	130	140	150	160
20		130	140	150	160	170
	orf83.pep	180	190			
	orf83a	180	190			
25		190	200	210	220	230
	orf83a	190	200	210	220	230
		240				

The complete length ORF83a nucleotide sequence <SEQ ID 315> is:

30	1	ATGAAAACCC	TGCTCNTCCT	CATCCCCCTC	GTCCTCACAG	CCTGCGGCAC
	51	ACTGACCGGC	ATACCCGCCC	ACGGCGGCGG	CAAACGCTTT	GCCGTCGAAC
	101	AAGAACTCGT	CGCCGCATCG	TCCCGCGCCG	CCGTCAAAGA	AATGGACTTG
	151	TCCGCCCTGA	AAGGACGCAA	AGCCGCCCTT	TACGTCTCCG	TTATGGGCGA
	201	CCAAGGTTTC	GGCAACATAA	GCGGCGGACG	CTACTCTATC	GACGCACGTA
	251	TACGCGGCGG	CTACCACAAC	AACCCCGAAA	GTGCCACCCA	ATACAGCTAC
35	301	CCCGCCTACG	ACACTACCGC	CACCACCAA	TCCGACGCGC	TCTCCAGCGT
	351	AACCACTTCC	ACATCGCTTT	TGAACGCCCC	CGCCGCGGCC	CTGACGAAAA
	401	ACAGCGGACG	CAAAGGCGAA	CGCTCCGCCG	GACTGTCCGT	CAACGGCAGC
	451	GGCGACTACC	GCAACGAAAC	CCTGCTCGCC	AACCCCGCG	ACGTTTCCTT
	501	CCTGACCAAC	CTCATCCAAA	CCGTCTTCTA	CCTGCGCGGC	ATCGAAGTCG
40	551	TACCGCCCGA	ATACGCGGAC	ACCGACGTAT	TCGTAACCGT	CGACGTATTC
	601	GGCACCCTCC	GCAGCCGCAC	CGAACTGCAC	CTCTACAACG	CCGAAACCTT
	651	TAAAGCCCAA	ACCAAGCTCG	AATATTTCGC	CGTTGACCGC	GACAGCCGGA
	701	AACTGCTGAT	TGCCCCATAA	ACCGCCGCCT	ACGAATCCCA	ATACCAAGAA
	751	CAATACGCCC	TCTGGATGGG	ACCTTACAGC	GTGCGCAAAA	CCGTCAAAGC
45	801	CTCAGACCGC	CTGATGGTCG	ATTTCTCCGA	CATCACCCTC	TACGGCGACA
	851	CAACCGCCCA	AAACCGTCCC	GACTTCAAAC	AAAACAACGG	TAAAAAACCC
	901	GATGTCGGCA	ACGAAGTCAT	CCGCCGCCGC	AAAGGAGGAT	AA

This encodes a protein having amino acid sequence <SEQ ID 316>:

50	1	MKTLXLPLVLT	ACGTLTGIPAHGGGKRFAVEQELVAAS	SRAAVKEMDL
	51	SALKGRKAAL	YVSVMGDQGS	GNISGGGRYSI
	101	PAYDTTATTK	SDALSSVTTS	TSLLNAPAAA
	151	GDYRNETLLA	NPRDVSFLTN	LIQTVFYLRG
	201	GTVRSRTELH	LYNAETLKAQ	TKLEYFAVDR
	251	QYALWMGPYS	VGKTVKASDR	LMVDFSDITP
55	301	DVGNEVIRRR	KGG*	

ORF83a and ORF83-1 show 98.4% identity in 313 aa overlap:

	10	20	30	40	50	60
orf83a.pep	MKTLXLPLVLT	ACGTLTGIPAHGGGKRFAVEQELVAASSRAAVKEMDLSALKGRKAAL				



[illegible]

35 Homology with a predicted ORF from *N.gonorrhoeae*

ORF83 shows 94.9% identity over a 197aa overlap with a predicted ORF (ORF83.ng) from *N. gonorrhoeae*:

40	orf83.pep	TLLLFIPLVLTXCGTLTGILAHGGGKRFAVEQELVAASSRAAVKEMDLSALKGRKAAX	58
	orf83ng	MKTLLLLLIPLVLTAAGTLTGIPAHGGGKRFAVEQELVAASSRAAVKEMDLSALKGRKAAL	60
45	orf83.pep	YVSVMGDQGSNGNISGGRYSIDALIRGGYHNNPESATQYSYPAYDTTATTKSDALSSVTTS	118
	orf83ng	YVSVMGDQGSNGNISGGRYSIDALIRGGYHNNPD SATRYSYPAYDTTATTKSDALSGVTTS	120
50	orf83.pep	TSLLNAPAAXLTKNSGRKGERSAGLSVNGTGDYRNETLLANPRDVSFLTNIQTVFYLRG	178
	orf83ng	TSLLNAPAAALTKNNGRKGERSAGLSVNGTGDYRNETLLANPRDVSFLTNIQTVFYLRG	180
50	orf83.pep	IEVVPXPYADTDVFTVDV	197
	orf83ng	IEVVPPEYADTDVFTVDVFGTVRSRTELHLYNAETLKAOTKLEYFAVDRDSRKLIIAPK	240

The complete length ORF83ng nucleotide sequence <SEQ ID 317> is:

55	1	ATGAAACCC	TGCTCTCT	CATCCCCCTC	GTA CTACCG	CCTGCGGCAC
	51	ACTGACCGC	ATACCGGCC	ACGGCGCGCG	CAAACGTTT	GCCGTGCAAT
	101	AGGAACCTGT	CGCCGATCG	TCCCGCGCGG	CCGTCAAAG	AATGGACTTG
	151	TCCGCCCTGA	AGGACGCAA	AGCCGCCCTT	TACGTCTCG	TTATGGGCGA
60	201	CCAAGTTTC	GGCAACATA	GCGGCGGACG	CTACTCCATC	GACGCACTGA
	251	TACGCGCGCG	CTACCACAAC	AACCCCGACA	GCGCCACCCG	ATACAGCTAC
	301	CCCGCTATG	ACATACCCG	CACCACCAA	TCCGACGCGC	TCTCCGGCGT
	351	AACCACTTCC	ACATCGCTTT	TGAACGCCCC	CGCCGCGCCG	CTCAGCAAAA
	401	ACAACGAGC	CAAAGGCGAA	CGCTCCGCCG	GACTGTCCGT	CAACGGGCAG
	451	GGCGACTACC	GCAACGAAAC	CCTGCTCGCC	AACCTCCCGG	ACGTTTCTCT
	501	CCTGACCAAC	CTCATCCAAA	CCGTCTTCTA	CTCGCGCGCG	ATCGAATCTG

551 TACCGCCCGA ATACGCCGAC ACCGACGTAT TCGTAACCGT CGACGTATTC  
 601 GGCACCGTCC GCAGCCGTAC CGAACTGCAC CTCTACAACG CCGAAACCCCT  
 651 TAAAGCCCAA ACCAAGCTCG AATATTTTCGC CGTCGACCGC GACAGCCGGA  
 701 AACTGCTGAT TGCCCCTAAA ACCGCCGCCT ACGAATCCCA ATACCAAGAA  
 751 CAATACGCCC TCTGGATGGG ACCTTACAGC GTCGGCAAAA CCGTCAAAGC  
 801 CTCAGACCGC CTGATGGTCG ATTTCTCCGA CATCACCCCC TACGGCGACA  
 851 CAACCGCCCA AAACCGTCCC GACTTCAAAC AAAACAACGG TAAAAACCCC  
 901 GATGTCGGCA ACGAAGTCAT CCGCCGCCGC AAAGGAGGAT AA

This encodes a protein having amino acid sequence <SEQ ID 318>:

10 1 MKTLLLLLIPL VLTACGTLTG IPAHHGGKRF AVEQELVAAS SRAAVKEMDL  
 51 SALKGRKAAL YVSVMGDQGS GNISGGRYSI DALIRGGYHN NPDSATRYSY  
 101 PAYDTTATTK SDALSGVTTT TSLNAPAAA LTKNNGRKGE RSAGLSVNGT  
 151 GDYRNETLLA NPRDVSFLTNI LIQTVFYLRG IEVVPPEYAD TDVFTVDFV  
 201 GTVRSRTELH LYNAETLKAQ TKLEYFAVDR DSRKLLIAPK TAAYESQYQE  
 15 251 QYALWMPYYS VGKTVKASDR LMVDFSDITP YGDTTAQNRP DFKQNNNGKNP  
 301 DVGNEVIRRR KGG\*

ORF83ng and ORF83-1 show 97.1% identity in 313 aa overlap

20 orf83-1.pep 10 20 30 40 50 60  
 MKTLLLLLIPLVLTACGTLTGIPAHHGGKRF AVEQELVAASSRAAVKEMDLSALKGRKAAL  
 orf83ng MKTLLLLLIPLVLTACGTLTGIPAHHGGKRF AVEQELVAASSRAAVKEMDLSALKGRKAAL  
 25 orf83-1.pep 70 80 90 100 110 120  
 YVSVMGDQGS GNISGGRYSIDALIRGGYHNNPESATQSY PAYDTTATTKSDALSSVTTT  
 orf83ng YVSVMGDQGS GNISGGRYSIDALIRGGYHNNPDSATRYSY PAYDTTATTKSDALSGVTTT  
 30 orf83-1.pep 130 140 150 160 170 180  
 TSLNAPAAALTKNNGRKGE RSAGLSVNGTGDYRNETLLANPRDVSFLTNIQTVFYLRG  
 orf83ng TSLNAPAAALTKNNGRKGE RSAGLSVNGTGDYRNETLLANPRDVSFLTNIQTVFYLRG  
 35 orf83-1.pep 190 200 210 220 230 240  
 IEVVPPEYADTDVFTVDFVGTGTVRSRTELHLYNAETLKAQTKLEYFAVDRDSRKLLITPK  
 orf83ng IEVVPPEYADTDVFTVDFVGTGTVRSRTELHLYNAETLKAQTKLEYFAVDRDSRKLLIAPK  
 40 orf83-1.pep 250 260 270 280 290 300  
 TAAYESQYQE QYALWMPYYS VGKTVKASDR LMVDFSDITP YGDTTAQNRP DFKQNNNGKPP  
 orf83ng TAAYESQYQE QYALWMPYYS VGKTVKASDR LMVDFSDITP YGDTTAQNRP DFKQNNNGKNP  
 45 orf83-1.pep 310  
 DVGNEVIRRRKGGX  
 orf83ng DVGNEVIRRRKGGX  
 310

Based on this analysis, including the presence of a putative ATP/GTP-binding site motif A (P-loop)  
 55 in the gonococcal protein (double-underlined) and a putative prokaryotic membrane lipoprotein  
 lipid attachment site (single-underlined), it is predicted that the proteins from *N.meningitidis* and  
*N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for  
 raising antibodies.

**Example 38**

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 319>:

```

1  ATGGCAGAGA TCTGTTTGAT AACCGGCACG CCCGGTTCAG GGAAACATT
5  51  AAAAATGGTT TCCATGATGG CGAATGATGA AATGTTTAAG CCTGATGAAA
101 AAGCCATACG CCGTAAAGTA TTTACGAACA TAAAGGCTT GAAATACCG
151 CACACCTACA TAGAAACGGA CGCAAAAAG CTGCCGAAAT CGACAGATGA
201 GCAGCTTTCG GCGCATGATA TGTACGAATG GATAAAGAAG CCCGAAAATA
251 TCGGGTCTAT TGTCAATTGA GATGAAGCTC AAGACGTATG GCCGGCACGC
10  301 TCGGCAGGTT CAAAAATCCC TGAAAATGTC CAATGGCTGA ATACGCACAG
351 ACATCAGGGC ATTGATATAT TTGTTTGTAC TCAAGGTCTT AAGCTTCTAG
401 ATCAAAATCT TAGAACGCTT GTACGGAAAC ATTACCACAT CGCTTCAAAC
451 AAGATGGGTA TCGGTACGCT TTTAGAATGG AAAATATGCG CGGACGATCC
501 CGTAAAAATG GCATCAAGCG CATTCTCCAG TATCTATACA CTGGATAAAA
15  551 AAGTTTATGA CTTGTATsrr TmmCGGGAAG TTCATACCGT AAATAAGGTC
601 AAGCGGTCAA AGTGGTTTTA CACTCTGCCa GTAATAGTAT TGCTGATTCC
651 CGTGTGTTGTC GGCCTGTCTT ATAAAATGTT GagCaGTTAC GGAAAAAACC
701 aGGAAGAACC CGCAGCACAA GAATCGGCGG CAACAGAACA GCAGGCAGTA
751 CTTCCGATA AAACAGAAGG CGAGCCGGTA AATAACGGCA ACCTTACCGC
20  801 AGATATGTTT GTTCCGACAT TGTCCGAaAA ACCCGrAAGC AAGCcgaTTT
851 ATAACGGTGT AAGGCAGGTA AGAACCTTTG AATATATAGC AGGCTGTATA
901 GAAGGCGGAA GAACCGGATG CGCCTGCTAT TCGCaTCAAG GGACGGCATT
951 gaAAGAAGTG ACGGaGTTGA TGTGcgaAgG aCTATGTaAA AAacGGCTTG
1001 CCGTTTAACC CaTACAAAGA AGAAAGCCAA GGGCAGGAAG TTCAGCAAAG
25  1051 CGCGCAGCAA CATTCCGACA GGGCGcCAAG TTGCCACATT GGGCGGAAAA
1101 CCGTAGCAGA ACCTAATGTA CGATAATTGG GAAGAACGCG GAAACCGTTT
1151 TGAAGGAATC GgCGGGGGC GTGGTCGGAT CGGCAAACTG A

```

This corresponds to the amino acid sequence <SEQ ID 320; ORF84>:

```

1  MAEICLITGT PGSGKTLKMV SMMANDEMFK PDEKAIRRKV FTNIKGLKIP
30  51  HTYIETDAKK LPKSTDEQLS AHDMYEWIKK PENIGSIVIV DEAQDVWPAR
101 SAGSKIPENV QWLNTHRHQG IDIFVLTQGP KLLDQNLRTL VRKHYHIASN
151 KMGMRITLLEW KICADDPVKM ASSAFSSIYT LDKKVYDLYX XAEVHTVNKV
201 KRSKWFTYTLF VIVLLIPVFV GLSYKMLSSY GKKQEEPAAQ ESAATEQQAV
251 LPDKTEGEPV NNGNLTADMV VPTLSEKPKS KPIYNGVRQV RTFEYIAGCI
35  301 EGGRTGCACY SHQGTALKEV TELMCKDYVK NGLPFNPYKE ESQGGVEQQS
351 AQQHSDRAQV ATLGGKPKQN LMYDNWEERG KPFEIGIGGV VGSAN*

```

Further work revealed the complete nucleotide sequence <SEQ ID 321>:

```

1  ATGGCAGAGA TCTGTTTGAT AACCGGCACG CCCGGTTCAG GGAAACATT
40  51  AAAAATGGTT TCCATGATGG CGAATGATGA AATGTTTAAG CCTGATGAAA
101 ACGGCATACG CCGTAAAGTA TTTACGAACA TAAAGGCTT GAAATACCG
151 CACACCTACA TAGAAACGGA CGCAAAAAG CTGCCGAAAT CGACAGATGA
201 GCAGCTTTCG GCGCATGATA TGTACGAATG GATAAAGAAG CCCGAAAATA
251 TCGGGTCTAT TGTCAATTGA GATGAAGCTC AAGACGTATG GCCGGCACGC
301 TCGGCAGGTT CAAAAATCCC TGAAAATGTC CAATGGCTGA ATACGCACAG
45  351 ACATCAGGGC ATTGATATAT TTGTTTGTAC TCAAGGTCTT AAGCTTCTAG
401 ATCAAAATCT TAGAACGCTT GTACGGAAAC ATTACCACAT CGCTTCAAAC
451 AAGATGGGTA TCGGTACGCT TTTAGAATGG AAAATATGCG CGGACGATCC
501 CGTAAAAATG GCATCAAGCG CATTCTCCAG TATCTATACA CTGGATAAAA
55  551 AAGTTTATGA CTTGTACGAA TCAGCGGAAG TTCATACCGT AAATAAGGTC
601 AAGCGGTCAA AGTGGTTTTA CACTCTGCCa GTAATAGTAT TGCTGATTCC
651 CGTGTGTTGTC GGCCTGTCTT ATAAAATGTT GAGCAGTTAC GGAAAAAACC
701 AGGAAGAACC CGCAGCACAA GAATCGGCGG CAACAGAACA GCAGGCAGTA
751 CTTCCGATA AAACAGAAGG CGAGCCGGTA AATAACGGCA ACCTTACCGC
801 AGATATGTTT GTTCCGACAT TGTCCGAaAA ACCCGAAAGC AAGCCGATTT
55  851 ATAACGGTGT AAGGCAGGTA AGAACCTTTG AATATATAGC AGGCTGTATA
901 GAAGGCGGAA GAACCGGATG CGCCTGCTAT TCGCaTCAAG GGACGGCATT
951 GAAAGAAGTG ACGGAGTTGA TGTGCAAGGA CTATGTAAAA AACGGCTTGC
1001 CGTTTAATCC ATACAAAGAA GAAAGCCAAG GGCAGGAAGT TCAGCAAAGC
60  1051 GCGCAGCAAC ATTCCGACAG GCGCAAGTT GCCACATTGG GCGGAAAACC
1101 GTAGCAGAAC CTAATGTACG ATAATGCGGA AGAACGCGG AAACCGTTTG
1151 AAGGAATCGG CGGGGGCGTG GTCGGATCGG CAACTGA

```

This corresponds to the amino acid sequence <SEQ ID 322; ORF84-1>:

1 MAEICLITGT PGSGKTLKMV SMMANDEMFK PDENGIRRKV FTNIKGLKIP  
 51 HTYIETDAKK LPKSTDEQLS AHDMYEWIKK PENIGSIVIV DEAQDVWPAR  
 101 SAGSKIPENV QWLNTHRHQG IDIFVLTQGP KLLDQNLRTL VRKHYHIASN  
 151 KMGMRITLLEW KICADDPVKM ASSAFSSIYT LDKKVYDLYE SAEVHTVNVK  
 201 KRKWFYTLF VIVLLIPVVF GLSYKMLSSY GKKQEEPAQ ESAATEQQAV  
 251 LPDKTEGEPV NNGNLTADMV VPTLSEKPES KPIYNGVRQV RTFEYIAGCI  
 301 EGGRTGCACY SHQGTALKEV TELMCKDYVK NGLPFNPYKE ESQGEVQQS  
 351 AQQHSRAQV ATLGKPK\*QN LMYDNWEERG KPFEIGGGV VGSAN\*

10 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF84 shows 93.9% identity over a 395aa overlap with an ORF (ORF84a) from strain A of *N. meningitidis*:

15	orf84.pep	MAEICLITGT PGSGKTLKMV SMMANDEMFK PDENGIRRKV FTNIKGLKIP HTYIETDAKK
	orf84a	MAEICLITGT PGSGKTLKMV SMMANDEMFK PDENGIRRKV FTNIKGLKIP HTYIETDAKK
20	orf84.pep	LPKSTDEQLS AHDMYEWIKK PENIGSIVIV DEAQDVWPAR SAGSKIPENV QWLNTHRHQG
	orf84a	LPKSTDEQLS AHDMYEWIKK PENIGSIVIV DEAQDVWPAR SAGSKIPENV QWLNTHRHQG
25	orf84.pep	IDIFVLTQGP KLLDQNLRTL VRKHYHIASN KMGMRITLLEW KICADDPVKM ASSAFSSIYT
	orf84a	IDIFVLTQGS KLLDQNLRTL VRKHYHIASN KMGMRITLLEW KICADDPVKM ASSAFSSIYT
30	orf84.pep	LDKKVYDLYE SAEVHTVNVK VRKSKWFYTLF VIVLLIPVVF GLSYKMLSSY GKKQEEPAQ
	orf84a	LDKKVYDLYE SAEVHTVNVK VRKSKWFYTLF VIVLLIPVVF GLSYKMLSSY GKKQEEPAQ
35	orf84.pep	ESAATEQQAV LPDKTEGEPV NNGNLTADMV VPTLSEKPES KPIYNGVRQV RTFEYIAGCI
	orf84a	ESAATEHQAV FQDKTEGEPV NNGNLTADMV VPTLSEKPES KPIYNGVRQV RTFEYIAGCV
40	orf84.pep	EGGRTGCACY SHQGTALKEV TELMCKDYVK NGLPFNPYKE ESQGEVQQS AQQHSRAQV
	orf84a	EGGRTGCTCY SHQGTALKEIT KEMCKDYARNGL PFNPYKEES QGRDVQQSEQHHSRDPQV
45	orf84.pep	ATLGKPKXQN LMYDNWEERG KPFEIGGGV VGSANX
	orf84a	ATLGKPKWQN LMYDNWQERG KPFEIGGGV VGSANX

55 The complete length ORF84a nucleotide sequence <SEQ ID 323> is:

1 ATGGCAGAGA TCTGTTTGAT AACCGGCACG CCCGGTTCAG GGAAACATT  
 51 AAAAATGGTT TCCATGATGG CAAACGATGA AATGTTTAAG CCGGATGAAA  
 101 ACGGCATACG CCGTAAAGTA TTTACGAACA TCAAAGGCTT GAAGATACCG  
 151 CACACCTACA TAGAAACGGA CGCGAAAAAG CTGCCGAAAT CGACAGATGA  
 201 GCAGCTTTCG GCGCATGATA TGTACGAATG GATAAAGAAG CCCGAAAATA

5  
10  
15  
20

```
251 TCGGGTCTAT TGTCATTGTA GATGAAGCTC AAGACGTATG GCCGGCACGC
301 TCGGCAGGTT CAAAAATCCC TGAAAATGTC CAATGGCTGA ATACGCACAG
351 ACATCAGGGC ATTGATATAT TTGTTTGGAC TCAAGGCTCT AAGCTTCTAG
401 ATCAAAATCT TAGAACGCTT GTACGGAAAC ATTACCACAT CGCTTCAAAC
451 AAGATGGGTA TGCCTACGCT TTTAGAATGG AAAATATGCG CGGACGATCC
501 CGTAAAAATG GCATCAAGCG CATTCTCCAG TATCTATACA CTGGATAAAA
551 AAGTTTATGA CTTGTACGAA TCAGCGGAAG TTCATACCGT AAATAAGGTC
601 AAGCGGTCAA AATGGTTTTA TACTCTGCCA GTAATAATAT TGCTGATTCC
651 CGTTTTTGTC GGCCTGTCTT ATAAAATGTT AAGTAGTTAT GGAAAAAAC
701 AGGAAGAACC CGCAGCACAA GAATCGGCGG CAACAGAACA TCAGGCAGTA
751 TTTCAGGATA AAACAGAAGG CGAGCCGGTA AACACGGTA ACCTTACCGC
801 AGATATGTTT GTTCCGACAT TGTCCGAAAA ACCCGAAAGC AAGCCGATTT
851 ATAACGGTGT AAGGCAGGTA AGAACCTTTG AATATATAGC AGGCTGTGTA
901 GAAGGCGGAA GAACCGGATG CACATGCTAT TCGCATCAAG GGACGGCATT
951 GAAAGAAATT ACAAAGGAAA TGTGCAAGGA TTACGCAAGA AACGGATTGC
1001 CGTTTAACCC ATATAAGAA GAAAGCCAAG GGCGGGATGT CCAGCAAAGT
1051 GAGCAGCACC ATTCGGACAG ACCGCAAGTT GCCACGTTGG GCGGAAAGCC
1101 GTGGCAAAT CTATGTATG ATAATTGGCA GGAGCGCGGA AAACCGTTTG
1151 AAGGAATCGG CGGGGGCGTG GTCGGATCGG CAAACTGA
```

20 This encodes a protein having amino acid sequence <SEQ ID 324>:

25

```
1 MAEICLITGT PGSGKTLKMV SMMANDEMFK PDENGIRRKV FTNIKGLKIP
51 HTYIETDAKK LPKSTDEQLS AHDMYEWIKK PENIGSIVIV DEAQDVWPAR
101 SAGSKIPENV QWLNTHRHQG IDIFVLTQGS KLLDQNLRTL VRKHYHIASN
151 KMGMRITLLEW KICADDPVKM ASSAFSSIYT LDKKVYDLYE SAEVHTVNVK
201 KRKSKWFYTLF VIILLIPVFLV GLSYKMLSSY GKKQEEPAQA ESAATEHQAV
251 FQDKTEGEPV NNGNLTADMF VPTLSEKPES KPIYNGVRQV RTEFYIAGCV
301 EGGRTGCTCY SHQGTALKEI TKEMCKDYAR NGLPFNPYKE ESQGRDVQQS
351 EQHHSRDPQV ATLGGKPWQN LMYDNWQERG KPFEGIGGV VGSAN*
```

ORF84a and ORF84-1 show 95.2% identity in 395 aa overlap:

30  
35  
40  
45  
50  
55  
60  
65

```
10 20 30 40 50 60
orf84a.pep MAEICLITGT PGSGKTLKMV SMMANDEMFK PDENGIRRKV FTNIKGLKIP HTYIETDAKK
orf84-1 MAEICLITGT PGSGKTLKMV SMMANDEMFK PDENGIRRKV FTNIKGLKIP HTYIETDAKK
10 20 30 40 50 60
orf84a.pep 70 80 90 100 110 120
LPKSTDEQLS AHDMYEWIKK PENIGSIVIV DEAQDVWPAR SAGSKIPENV QWLNTHRHQG
orf84-1 LPKSTDEQLS AHDMYEWIKK PENIGSIVIV DEAQDVWPAR SAGSKIPENV QWLNTHRHQG
70 80 90 100 110 120
orf84a.pep 130 140 150 160 170 180
IDIFVLTQGS KLLDQNLRTL VRKHYHIASN KMGMRITLLEW KICADDPVKM ASSAFSSIYT
orf84-1 IDIFVLTQGS KLLDQNLRTL VRKHYHIASN KMGMRITLLEW KICADDPVKM ASSAFSSIYT
130 140 150 160 170 180
orf84a.pep 190 200 210 220 230 240
LDKKVYDLYE SAEVHTVNVK VRKSKWFYTLF VIILLIPVFLV GLSYKMLSSY GKKQEEPAQA
orf84-1 LDKKVYDLYE SAEVHTVNVK VRKSKWFYTLF VIILLIPVFLV GLSYKMLSSY GKKQEEPAQA
190 200 210 220 230 240
orf84a.pep 250 260 270 280 290 300
ESAATEHQAV FQDKTEGEPV NNGNLTADMF VPTLSEKPES KPIYNGVRQV RTEFYIAGCV
orf84-1 ESAATEHQAV FQDKTEGEPV NNGNLTADMF VPTLSEKPES KPIYNGVRQV RTEFYIAGCI
250 260 270 280 290 300
orf84a.pep 310 320 330 340 350 360
EGGRTGCTCY SHQGTALKEI TKEMCKDYAR NGLPFNPYKE ESQGRDVQQS EQHHSRDPQV
orf84-1 EGGRTGCTCY SHQGTALKEI TKEMCKDYAR NGLPFNPYKE ESQGRDVQQS EQHHSRDPQV
310 320 330 340 350 360
370 380 390
```

```

orf84a.pep    ATLGGKPWQNLMYDNWQERGKPFEGIGGGVVGGSANX
              |||||:|||||
orf84-1       ATLGGKPXQNLMYDNWEERGKPFEGIGGGVVGGSANX
              370      380      390

```

5

Homology with a predicted ORF from *N.gonorrhoeae*

ORF84 shows 94.2% identity over a 395aa overlap with a predicted ORF (ORF84.ng) from *N. gonorrhoeae*:

10	orf84.pep	MAEICLITGTGPGSGKTLKMSVSMANDEMFKPDEKAIRRKVFTNIKGLKIPHTYIETDAKK      :     :     :	60
	orf84.ng	MAEICLITGTGPGSGKTLKMSVSMANDEMFKPDENGVRKRVFTNIKGLKIPHTHIETDAKK	60
15	orf84.pep	LPKSTDEQLSAHDMYEWIKKPENIGSIVIVDEAQDVWPARSAGSKIPENVQWLNTHRHQG      :     :     :	120
	orf84.ng	LPKSTDEQLSAHDMYEWIKKPENVGAIIVIVDEAQDVWPARSAGSKIPENVQWLNTHRHQG	120
20	orf84.pep	IDIFVLTQGPBKLLDQNLRTLVRKHYYHIAASNKMGMRTLLEWKICADDPVKMASSAFSSIIYT      :     :     :	180
	orf84.ng	IDIFVLTQGPBKLLDQNLRTLVRKHYYHIAANKMGLRTLLEWKVCADDPVKMASSAFSSIIYT	180
25	orf84.pep	LDKKVYDLYXXAEVHTVNKVKRSKWFTLPVIVLLIPVFVGLSYKMLSSYGKKQEEPAAQ      :     :     :	240
	orf84.ng	LDKKVYDLYESAEIHTVNKVKRSKWFYALPVIIILLPLFVGLSYKMLGSYGKKQEEPAAQ	240
30	orf84.pep	ESAAATEQQAVLPDKTEGEPVNNGNLTADMFVPTLSEKXPXSKPIYNGVRQVRTFEYIAGCI      :     :     :	300
	orf84.ng	ESAAATEQQAVLPDKTEGESVNNGNLTADMFVPTLPEKPESKPIYNGVRQVRTFEYIAGCI	300
35	orf84.pep	EGGRTGCACYSHQGTALKEVTELMCKDYVKNGLPFNPYKEESQGQEVQSAQQHSDDRAQV      :     :     :	360
	orf84.ng	EGGRTGCTCYSHQGTALKEVTELMCKDYVKNGLPFNPYKEESQGQEVQSAQQHSDDRAQV	360
40	orf84.pep	ATLGGKFPXQNLMYDNWEERGKPFEGIGGGVVGGSAN 395      :     :     :	420
	orf84.ng	ATLGGKPPQNLMYDNWEERGKPFEGIGGGVVGGSAN 395	420

The complete length ORF84ng nucleotide sequence <SEQ ID 325> is:

	1	ATGGCAGAAA	TCTGTTTGAT	AACCGGCACG	CCCGGTTTCAG	GGAAAAACATT
	51	AAAAATGGTT	TCCATGATGG	CAACGATGA	AATGTTTAAG	CCAGATGAAA
40	101	ACGGCGTACG	CCGTAAAGTA	TTTACGAACA	TCAAAGGTTT	GAAAGTACCG
	151	CACACCCACA	TAGAAAACAGA	CGCAAAGAAG	GTCCCGAAT	CAACCGATGA
	201	ACAGCTTTTCG	CGCGATGATA	TGTATGAATG	CATCAAGAAG	CCTGAAAacg
	251	tcggcgCAAT	CGTTATTGTC	GATGAGGCGC	AAGACGTATG	GCCCGCACGC
	301	TccgCAGGTT	CGAAAATCCC	CGAAAACGTC	CAATGGCTGA	ACACACACAG
45	351	GCATCAGGCG	ATAGATATAT	TTGTATTGAC	ACAAGGTCCT	AAACTCTTAG
	401	ATCAGAACTT	CGCAACATTG	GTTAAAAGAC	ATTACCACAT	TGCGGCCAAC
	451	AAAATGGGTT	TGCGTACCTT	GCTTGAATGG	AAAGTATGCG	CGGATGACCC
	501	GGTAAAAATG	GCATCAAGTG	CATTTTCCAG	TATCTACACA	CTGGATAAAA
	551	AAGTTTATGA	CTTGTAACGA	TCCGCAGAAA	TTCCACCGGT	AAACAAAGTC
50	601	AAGCGTTCAA	AATGGTTTTA	TGCATTGCCC	GTCATCATAT	TATTGATTCC
	651	GCTATTTGTC	GGTTTGTCTT	ACAAAATGTT	GGGCAGTTAC	GGAAAAAACC
	701	AGGAAGAACC	CGCAGCACAA	GAATCGGCGG	CACAGAACA	CGAGGCAGTA
	751	CTTCCGGAATA	AAACAGAAGG	AGAATCGGTG	AATAACGGAA	ACCTTACCGC
	801	AGATATGTTT	GTTCCGACAT	TGCCCGAAAA	ACCCGAAAGC	AAGCCGATTT
55	851	ATAACGGTGT	AAGGCAGGTA	AGGACCTTTG	AATATATAGC	AGGCTGTATA
	901	GAAAGCGGAA	GAACCGGATG	CACCTGCTAT	TCGCATCAAG	GGACGGCATT
	951	GAAAGAAGTG	ACGAGTTTGA	TGTGCAAGGA	CTATGTA AAA	ACACGGCTTGC
	1001	CGTTTAAACC	ATACAAAGAA	GAAAGCCAAG	GGCAGGAAGT	TCAGCAAAGC
	1051	GCGCAGCAAC	ATTCCGAGAT	GGCGCAAGTT	GCCACCTTGG	GCGGAAAAACC
60	1101	CGACGAGAAC	CTAATTGTACG	ACAATTGGGA	AGAACCGCGG	AAACCGTTTG
	1151	AAGGAATCGG	CGGGGGCGTG	GTCCGATCCG	CAAACTGA	

This encodes a protein having amino acid sequence <SEO ID 326>:

1 MAEICLITGT PGSGKTLKMV SMMANDEMFK PDENGVRKVF FTNIKGLKIP

51 HTHIETDAKK LPKSTDEQLS AHDMEYEWIKK PENVGAIVIV DEAQDVWPAR  
 101 SAGSKIPENV QWLNTHRHQG IDIFVLTQGP KLLDQNLRTL VKRHYHIAAN  
 151 KMGLRTLLEW KVCADDPVKM ASSAFSSIYT LDKKVYDLYE SAEIHTVNVK  
 201 KRSKWYFALP VIILLIPLFV GLSYKMLGSY GKKQEEPAQAQ ESAATEQQAV  
 251 LPDKTEGESV NNGNLTADMF VPTLPEKPES KPIYNGVRQV RTFEYIAGCI  
 301 EGGRTGCTCY SHQGTALKEV TELMCKDYVK NGLPFNPYKE ESQGQEVQQS  
 351 AQQHSDRAQV ATLGGKPQQN LMYDNWEERG KPFEGIGGGV VGSAN\*

ORF84ng and ORF84-1 show 95.4% identity in 395 aa overlap:

10	orf84-1.pep	10	20	30	40	50	60
	MAEICLITGTPGSGKTLKMVSMMANDEMFKPDENGIRRKVFTNIKGLKIPHTYIETDAKK						
	orf84ng	10	20	30	40	50	60
	MAEICLITGTPGSGKTLKMVSMMANDEMFKPDENGVRKVFTNIKGLKIPHTHIETDAKK						
15	orf84-1.pep	70	80	90	100	110	120
	LPKSTDEQLSAHDMYEWIKKPENIGSIVIVDEAQDVWPARSAGSKIPENVQWLNTHRHQG						
	orf84ng	70	80	90	100	110	120
	LPKSTDEQLSAHDMYEWIKKPENVGAIVIVDEAQDVWPARSAGSKIPENVQWLNTHRHQG						
20	orf84-1.pep	130	140	150	160	170	180
	IDIFVLTQGP KLLDQNLRTL VKRHYHIA SNKMG MRTLLEWK ICADDPVK MASSAFSSIYT						
	orf84ng	130	140	150	160	170	180
	IDIFVLTQGP KLLDQNLRTL VKRHYHIA ANKMGLRTLLEWKVCADDPVKMASSAFSSIYT						
25	orf84-1.pep	190	200	210	220	230	240
	LDKKVYDLYE SAEVHTVNVKVRKSKW FYTLPVIVLLIPV FVGLSYKMLSSYGKKQEEPAQAQ						
	orf84ng	190	200	210	220	230	240
	LDKKVYDLYE SAEIHTVNVKVRKSKW FVYALPVIILLIPLFVGLSYKMLGSYGKKQEEPAQAQ						
30	orf84-1.pep	250	260	270	280	290	300
	ESAATEQQAVLPDKTEGE PVNNGNLTADMFVPTLSEKPESKPIYNGVRQVRTFEYIAGCI						
	orf84ng	250	260	270	280	290	300
	ESAATEQQAVLPDKTEGESVNNGNLTADMFVPTLPEKPESKPIYNGVRQVRTFEYIAGCI						
35	orf84-1.pep	310	320	330	340	350	360
	EGGRTGCACYS HQGTALKEV TELMCKDYVK NGLPFNPYKEESQGQEVQQSAQQHSDRAQV						
	orf84ng	310	320	330	340	350	360
	EGGRTGCTCYS HQGTALKEV TELMCKDYVK NGLPFNPYKEESQGQEVQQSAQQHSDRAQV						
40	orf84-1.pep	370	380	390			
	ATLGGKXPQNLMYDNWEERGKPFEGIGGGVVGSANX						
	orf84ng	370	380	390			
	ATLGGKPPQQNLMYDNWEERGKPFEGIGGGVVGSANX						

50 Based on this analysis, including the presence of a putative transmembrane domain (single-underlined) in the gonococcal protein, and a putative ATP/GTP-binding site motif A (P-loop, double-underlined), it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 39

55 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 327>:

1 GTGGT TTTCC TGAATGCCGA CAACGGGATA TTGGTTCAGG ACTGCCTTT  
 51 TGAAGTCAAA CTGAAAAAAT TCCATATCGA TTTTACAAAT ACGGGTATGC  
 101 CGCGTGATTT CGCCAGCGAT ATTGAAGTGA CGGACAAGGC AACCGGTGAG

151 AACTCGAGC GCACCATCCG CGTGAACCAT CCTTTGACCT TGCACGGCAT  
 201 CACGATTTAT CAGGCGAGTT TTGCCGACGG CGGTTCCGAT TTGACATTCA  
 251 AGGCGTGGAA TTTGGGTGAT GCTTCGCGCG AGCCTGTCGT GTTGAAGGCA  
 301 ACATCCATAC ACCAGTTTCC GTTGAAATT GGCAAACACA AATATCGTCT  
 351 TGAGTTCGAT CAGTTCACTT CTATGAATGT GGAGGACATG AGCGAGGGCG  
 401 CGGAACGGGA AAAAAGCCTG AAATCCACGC TGCCCGATGT CCGCGCCGTT  
 451 ACTCAGGAAG GTCACAAATA CACCAAT... ..TACCG  
 501 TATCCGTGAT GCGCCAGGCC AGGCGGTGCA ATATAAAAAC TATATGCTGC  
 551 CGGTTTTCGA GGAACAGGAT TATTTTGGGA TTACCGGCAC GCGCAGCGC.  
 601 TTGCAGCAGC AATACCGCTG GCTGCGTATC CCCTTGGACA AGCAGTTGAA  
 651 AGCGGACACC TTTATGGCAT TCGGTGAGTT TTGAAAAGAT GGGGAAGGGC  
 701 GCAAACGTCT .GTTGCCGAC GCAACCAAAG GCGCACCTGC CGAAATCCGC  
 751 GAACAATTCA TGCTGGCTGC GGAACACACG CTGAACATCT TTGCACAAAA  
 801 AGGCTATTTG GGATTGGACG AATTTATTAC GTCCAATATC CCGAAAGAGC  
 851 AGCAGGATAA GATGCAGGGC TATTTCTACG AAATGCTTTA CCGCGTGATG  
 901 AACGCTGCTT TGGATGAAAC CAT.ACCCGG TACGGCTTGC CCGAATGGCA  
 951 GCAGGATGAA GCGCGGAATC GTTTCCTGCT GCACAGTATG GATGCGTACA  
 1001 CGGGTTTGAC CGAATATCCC GCGCCTATGC TGCTGCAACT TGATGGGTTT  
 1051 TCCGAGGTGC GTTCGTGCGG TTTGCAGATG ACCCGTTCCC C.GGTCCGCT  
 1101 TTTGCTCTAT CTC...

This corresponds to the amino acid sequence <SEQ ID 328; ORF88>:

1 MVFLNADNGI LVQDLPEVK LKKFHIDFYN TGMPRDFASD IEVTDKATGE  
 51 KLERTIRVNH PLTLHGITY QASFADGGSD LTFKAWNLGD ASREPVLVKA  
 101 TSIHQFPLEI GKHKYRLEFD QFTSMNVEDM SEGAEREKSL KSTLPDVRVAV  
 151 TQEGHKYTNX XXXXXYRIRD APGQAVEYKN YMLPVLQEQD YFWITGTRSX  
 201 LQQQYRWLRI PLDKQLKADT FMALREFLKD GEGRKRXVAD ATKGAPAEIR  
 251 EQFMLAAENT LNIFAQKGYL GLDEFITSNI PKEQQDKMQG YFYEMLYGVM  
 301 NAALDETXTX YGLPEWQQDE ARNRFLHLSM DAYTGLTEYP APMLLQLDGF  
 351 SEVRSSGLQM TRSXGPLLVY L...

Further work revealed the complete nucleotide sequence <SEQ ID 329>:

1 ATGAGTAAAT CCCGTAGATC TCCCCCACTT CTTTCCCGTC CGTGTTTCGC  
 51 TTTTTCAGC TCCATCGCT TTGCAGTCGC TTTGCTCAGT CTGCTGGGTA  
 101 TTGCATCGGT TATCGGTACG GTGTGCAGC AAAACCAAGC GCACAGCGAT  
 151 TATTTGGTCA AATTCGGATC GTTTTGGGCG CAGATTTTTC GTTTTCTGGG  
 201 ACTGTATGAC GTCTATGCTT CCGCATGGTT TGCTGTTATC ATGATGTTTT  
 251 TGGTGGTTTC TACCAATTTG TGCTGATTC GCAATGTGCC GCCGTTCCTG  
 301 CGCGAAATGA AGTCTTTTCG GGAAGAGTT AAAGAAAAAT CTCTGCGGCG  
 351 GATGCGCCAT TCTTCGCTGT TGGATGTAAA AATTGCGCCC GAGGTGCGCA  
 401 AACGTTATCT GGAAGTACAA GGTTTTCAGG GAAAAACCAT TAACCGTGAA  
 451 GACGGGTCCG TTCTGATTGC CGCCAAAAAA GGCACAATGA ACAATGGGG  
 501 CTATATCTTT GCCCATGTTG CTTTGATTGT CATTTGCCTG GCGGGTTTGA  
 551 TAGACAGTAA CCTGCTGTTG AACTGGGTA TGCTGACCGG TCGGATTGTT  
 601 CCGGACAATC AGGCGGTTTA TGCCAAGGAT TTCAAGCCCG AAAGTATTTT  
 651 GGGTGCCTCC AATCTCTCAT TTAGGGGCAA CGTCAATATT TCCGAGGGCG  
 701 AGAGTGCCTT TTGAAGTCAA ACTGAAAAAA TTCCATATCG ATTTTACAA  
 751 GACTTGCTT TGAAGTCAA ACTGAAAAAA TTCCATATCG ATTTTACAA  
 801 TACGGGTATG CCGCGTGATT TCGCCAGCGA TATTGAAGTG ACGGACAAGG  
 851 CAACCGGTGA GAACTCGAG CGCACCATCC GCGTGAACCA TCCTTTGACC  
 901 TTGCACGGCA TCACGATTTA TCAGGCGAGT TTTGCCGACG GCGGTTCCGA  
 951 TTTGACATTC AAGGCGTGGA ATTTGGGTGA TGCTTCGCGC GAGCCTGTGC  
 1001 TGTTGAAGGC AACATCCATA CACCAATTTT CGTTGGAAAT TGGCAACAC  
 1051 AAATATCGTC TTGAGTTTGA TCAGTTCACT TCTATGAATG TGGAGGACAT  
 1101 GAGCGAGGGC GCGGAACGGG AAAAAGCCT GAAATCCACG CTGAAAGATG  
 1151 TCCGCGCCGT TACTCAGGAA GGTAAAAAAT ACACCAATAT CCGCCCTTCC  
 1201 ATTGTTTACC GTATCCGTGA TGCGGCAGGG CAGGCGGTGC AATATAAAAA  
 1251 CTATATGCTG CCGGTTTTGC AGGAACAGGA TTATTTTGG ATTACCGGCA  
 1301 CGCGCAGCGG CTGTCAGCAG CAATACCGCT GGCTGCGTAT CCCCTTGGAC  
 1351 AAGCAGTTGA AAGCGGACAC CTTTATGGCA TTGCGTGAGT TTTTGAAGA  
 1401 TGGGGAAGGG CGCAACGTC TGGTTCGCGA CGCAACCAAA GCGCAGCTG  
 1451 CCGAAATCCG CGAACAATTC ATGCTGGCTG CGGAAACAC GCTGAACATC  
 1501 TTTGCACAAA AAGGCTATTT GGGATTGGAC GAATTTATTA CGTCCAAAT  
 1551 CCGGAAAGAG CAGCAGGATA AGATGCAGG CTATTCTAC GAAATGCTTT  
 1601 ACGGCGTGAT GAACGCTGCT TTGGATGAAA CCATACGCGG GTACGGCTTG  
 1651 CCCGAATGGC AGCAGGATGA AGCGCGGAAT CGTTTCCTGC TGCACAGTAT  
 1701 GGATGCGTAC ACGGGTTTGA CCGAATATCC CGCGCCTATG CTGCTGCAAC  
 1751 TTGATGGGTT TTCCGAGGTG CGTTCGTCGG GTTTCAGAT GACCCGTTCC



5  
1801 CCGGGTGGCG TTTTGGTCTA TCTCGGCTCG GTGCTGTTGG TATTGGGTAC  
1851 GGTATTGATG TTTTATGTGC GCGAAAAACG GGCGTGGGTA TTGTTTTTCTAG  
1901 ACGGCAAAAT CCGTTTTTGC ATGTCTTCGG CCCGCAGCGA ACGGGATTG  
1951 CAGAAGGAAT TTCCAAAACA CGTCGAGAGT CTGCAACGGC TCGGCAAGGA  
2001 CTTGAATCAT GACTGA

This corresponds to the amino acid sequence <SEQ ID 330; ORF88-1>:

1 MSKSRSPPL LSRPWFAFFS SMRFAVALLS LLGIASVIGT VLQONQPQTD  
51 YLVKFGSFWA QIFGFLGLYD VYASAWFVVI MMFLVVSTSL CLIRNVPPFW  
101 REMKSFREKV KEKSLAAMRH SSLLDVKIAP EVAKRYLEVQ GFQGKTINRE  
151 DGSVLIAAKK GTMNKGYIF AHVALIVICL GGLIDSNLLL KLGMTGRIV  
201 PDNQAVYAKD FKPESILGAS NLSFRGNVNI SEGQSADVVF LNADNGILVQ  
251 DLPFEVKLKK FHIDFYNTGM PRDFASDIEV TDKATGEKLE RTIRVNHPLT  
301 LHGITIYQAS FADGGSDLTF KAWNLDASR EPVVLKATSI HQFPLEIGHK  
351 KYRLEFDQFT SMNVEDMSEG AEREKSLKST LNDVRAVTQE GKKYTNIGPS  
15 401 IVYRIRDAAG QAVEYKNYML PVLQEQDYFW ITGTRSGLQQ QYRWLRIPLD  
451 KQLKADTFMA LREFLDGEG RKRLVADATK GAPAEIREQF MAAENTLNI  
501 FAQKGYLGLD EFITSNIPKE QQDKMQGYFY EMLYGMNAA LDETIRRYGL  
551 PEWQQDEARN RFLHSMDAY TGLTEYPAPM LLQLDGFSEV RSSGLQMTS  
601 PGALLVYLG S VLLVLGTVLM FYVREKRAWV LFSDGKIRFA MSSARSERDL  
20 651 QKEFPKHVES LQRLGKDLNH D\*

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF88 shows 95.7% identity over a 371aa overlap with an ORF (ORF88a) from strain A of *N. meningitidis*:

25 orf88.pep MVFLNADNGILVQDLDPFEVKLKKFHIDFYN  
orf88a AKDFKPESILGASNLSFRGNVNI SEGQSADVVF LNADNGILVQDLDPFEVKLKKFHIDFYN  
210 220 230 240 250 260  
30 orf88.pep 40 50 60 70 80 90  
TGMPRDFASDIEVTDKATGEKLERTIRVNHPLTLHGITIYQASFADGGSDLTFKAWNLD  
orf88a TGMPRDFASDIEVTDKATGEKLERTIRVNHPLTLHGITIYQASFADGGSDLTFKAWNLD  
270 280 290 300 310 320  
35 orf88.pep 100 110 120 130 140 150  
ASREPVLKATSIHQFPLEIGHKHKYRLEFDQFTSMNVEDMSEGAEREKSLKSTLPDVRV  
orf88a ASREPVLKATSIHQFPLEIGHKHKYRLEFDQFTSMNVEDMSEGAEREKSLKSTLNDVRV  
330 340 350 360 370 380  
40 orf88.pep 160 170 180 190 200 210  
TQEGHKYTNXXXXXXYRIRDAPQAVEYKNYMLPVLQEQDYFWITGTRSLQQQYRWLRI  
orf88a TQEGKKYTNIGPSIVYRIRDAAGQAVEYKNYMLPVLQEQDYFWITGTRSGLQQQYRWLRI  
390 400 410 420 430 440  
45 orf88.pep 220 230 240 250 260 270  
PLDKQLKADTFMALREFLDGEGRKRVADATKGAPAEIREQFMLAAENTLNIFAQKGYL  
orf88a PLDKQLKADTFMALREFLDGEGRKRLVADATKGAPAEIREQFMLAAENTLNIFAQKGYL  
450 460 470 480 490 500  
50 orf88.pep 280 290 300 310 320 330  
GLDEFITSNIPKEQQDKMQGYFYEMLYGMNAAALDETXYTRYGLPEWQQDEARNRFLHSM  
orf88a GLDEFITSNIPKEQQDKMQGYFYEMLYGMNAAALDETIRRYGLPEWQQDEARNRFLHSM  
510 520 530 540 550 560  
55 orf88.pep 340 350 360 370  
DAYTGLTEYPAPMLLQLDGFSEVRSSGLQMTSXXGPLLVYL

```

      |||
orf88a DAYTGLTEYPAPMLQLDGFSEVRSSGLQMTSPGALLVYLGSVLLVLGTVLMFYVREKR
      570      580      590      600      610      620

5    orf88a AWVLFSDGKIRFAMSSARSERDLQKEFPKHVESLQRLGKDLNHDH
      630      640      650      660      670

```

The complete length ORF88a nucleotide sequence <SEQ ID 331> is:

```

1  ATGAGTAAAT CCCGTAGATC TCCCCCACTT CTTTCCCGTC CGTGGTTCGC
51  TTTTTCAGC TCCATGCGCT TTGCGGTTCG TTTGCTCAGT CTGCTGGGTA
101 TTGCATCGGT TATCGGTACG GTTTGCAGC AAAACCAGCC GCAGACGGAT
151 TATTTGGTCA AATTCGGATC GTTTTGGGCG CAGATTTTGT GTTTTCTGGG
201 ACTGTATGAC GTCTATGCTT CGGCATGGTT TGTCGTTATC ATGATGTTTT
251 TGGTGGTTTC TACCAGTTTG TGCTGATTC GCAATGTGCC GCCGTTCTGG
301 CGCGAAATGA AGTCTTTTCG GGAAAAGGTT AAAGAAAAAT CTCTGGCGGC
351 GATGCGCCAT TCTTCGCTGT TGGATGTAAA AATGCGCCCC GAGGTTGCCA
401 AACGTTATCT GGAAGTACAA GGTTTTCAGG GAAAAACCAT TAACCGTGAA
451 GACGGGTTCG TTCTGATTGC CGCCAAAAAA GGCACAATGA ACAATGGGG
501 CTATATCTTT GCCCATGTTG CTTTGATTGT CATTTGCCCTG GCGGGTTGA
551 TAGACAGTAA CCTGCTGTGT AAAGTGGGTA TGCTGACCGG TCGGATTGTT
601 CCGGACAATC AGGCGGTTTA TGCCAAGGAT TTCAAGCCCG AAAGTATTTT
651 GGGTGCCTCC AATCTCTCAT TTAGGGGCAA CGTCAATATT TCCGAGGGGC
701 AGAGTGCGGA TGTGGTTTTT CTGAATGCCG ACAACGGGAT ATTGGTTCAG
751 GACTTGCCCTT TTGAAGTCAA ACTGAAAAAA TTCCATATCG ATTTTACAA
801 TACGGGTATG CCGCGCGATT TTGCCAGTGA TATTGAAGTA ACGGATAAGG
851 CAACCGGTGA GAAACTCGAG CGCACCATCC GCGTGAACCA TCCTTTGACC
901 TTGCACGGCA TCACGATTTA TCAGGCGAGT TTTGCCGACG GCGGTTCCGA
951 TTTGACATTC AAGGCGTGGA ATTTGGGTGA TGCTTCGCGC GAGCCTGTCC
1001 TGTGAAGGC AACATCCATA CACCAATTTC CGTTGGAAAT TGGCAAACAC
1051 AAATATCGTC TTGAGTTCGA TCAGTTTACT TCTATGAATG TGGAGGACAT
1101 GAGCGAGGGC GCGGAACGGG AAAAAAGCCT GAAATCCACG CTGAACGATG
1151 TCCGCGCCGT TACTCAGGAA GGTAAAAAAT ACACCAATAT CCGCCCTTCC
1201 ATTGTTTACC GTATCCGTGA TCGGCGAGGG CAGGCGGTGC AATATAAAAA
1251 CTATATGCTG CCGGTTTTGC AGGAACAGGA TTATTTTGG ATTACCGGCA
1301 CGCGCAGCGG CTTCGAGCAG CAATACCGCT GGCTGCGTAT CCCCTTGGAC
1351 AAGCAGTTGA AAGCGGACAC CTTTATGGCA TTGCGTGAGT TTTTGAAGA
1401 TGGGAAGGG CGCAACGTC TGGTTGCCGA CGCAACCAA GGCACCTG
1451 CCGAAATCCG CGAACAATTC ATGCTGGCTG CGGAAACAC GCTGAACATC
1501 TTTGCACAAA AAGGCTATTT GGGATTGGAC GAATTTATTA CGTCCAATAT
1551 CCCGAAAGAG CAGCAGGATA AGATGCAGGG CTATTTCTAC GAAATGCTTT
1601 ACGGCGTGAT GAACGCTGCT TTGGATGAAA CCATACGCCG GTACGGCTTG
1651 CCCGAATGCG AGCAGGATGA AGCGCGGAAT CGTTTCCTGC TGCACAGTAT
1701 GGATGCGTAC ACGGTTTGA CCGAATATCC CGCGCCTATG CTGTGCAAC
1751 TTGATGGGTT TCCGAGGTG CGTTCGTCGG GTTGCAGAT GACCCGTTCC
1801 CCGGGTGCGC TTTTGGTCTA TCTCGGCTCG GTGCTGTTGG TATGGGTAC
1851 GGTATTGATG TTTTATGTGC GCGAAAAACG GCGGTGGGTA TTGTTTTCAG
1901 ACGGCAAAAT CCGTTTTGCC ATGTCTTCGG CCCGCAGCGA ACGGATTTG
1951 CAGAAGGAAT TTCCAAACA CGTCGAGAGT CTGCAACGGC TCGGCAAGGA
2001 CTTGAATCAT GACTGA

```

This encodes a protein having amino acid sequence <SEQ ID 332>:

```

50  1  MSKSRSPPL LSRPWFAFFS SMRFAVALLS LLGIASVIGT VLQONQPQTD
      51  YLVKFGSFWA QIFGFLGLYD VYASAWFVVI MMFLVVSTSL CLIRNVPPFW
101  REMKSFREKV KEKSLAAMRH SSLLDVKIAP EVAKRYLEVQ GFQGKTINRE
151  DGSVLIAAKK GTMNKWGYIF AHVALIVICL GGLIDSNLLL KLGMLTGRIV
201  PDNQAVYAKD FKPEILGAS NLSFRGNVNI SEGQSADVVF LNAADNGILVQ
251  DLPFEVKLKK FHIDFYNTGM PRDFASDIEV TDKATGEKLE RTIRVNHPLT
301  LHGITIYQAS FADGGSDLTF KAWNLDASR EPVVLKATSI HQFPLEIGKH
351  KYRLEFDQFT SMNVEDMSEG AEREKSLKST LNDVRAVTQE GKKTYNIGPS
401  IVYRIRDAAG QAVEYKNYML PVLQEQDYFW ITGTRSGLQQ QYRWLRIPLD
451  KQLKADTFMA LREFLKDGEG RKRLVADATK GAPAEIREQF MLAAENTLNI
501  FAQKGYLGLD EFITSNIPKE QQDKMQGYFY EMLYGVMNAA LDETIRRYGL
551  PEWQQDEARN RFLHSMDAY TGLTEYPAPM LLQLDGFSEV RSSGLQMTS
601  PGALLVYLG S VLLVLGTVLM FYVREKRAWV LFS DGKIRFA MSSARSERDL
651  QKEFPKHVES LQRLGKDLNH D*

```

ORF88a and ORF88-1 100.0% identity in 671 aa overlap:

	orf88a.pep	MSKSRSPPLLSRPWF AFFSSMRFAVALLSLLGIASVIGTVLQQNQPTDYLKFGSFWA	60
	orf88-1	MSKSRSPPLLSRPWF AFFSSMRFAVALLSLLGIASVIGTVLQQNQPTDYLKFGSFWA	60
5	orf88a.pep	QIFGFLGLYDVYASAWFVIMMFLVVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH	120
	orf88-1	QIFGFLGLYDVYASAWFVIMMFLVVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH	120
10	orf88a.pep	SSLLDVKIAPEVAKRYLEVQGFQGKTINREDGSVLI AAKGTMNKWGYIFAHVALIVICL	180
	orf88-1	SSLLDVKIAPEVAKRYLEVQGFQGKTINREDGSVLI AAKGTMNKWGYIFAHVALIVICL	180
	orf88a.pep	GGLIDSNLLKLGM LTGRIVPDN QAVYAKDFKPESILGASNLSFRGNVNI SEGQSADVVF	240
15	orf88-1	GGLIDSNLLKLGM LTGRIVPDN QAVYAKDFKPESILGASNLSFRGNVNI SEGQSADVVF	240
	orf88a.pep	LNADNGILVQDL PFEVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLERTIRVNHPLT	300
	orf88-1	LNADNGILVQDL PFEVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLERTIRVNHPLT	300
20	orf88a.pep	LHGITIYQASFADGGSDLT FKA WNLGDASREP VVLKATSIHQFPLEIGKHKYRLEFDQFT	360
	orf88-1	LHGITIYQASFADGGSDLT FKA WNLGDASREP VVLKATSIHQFPLEIGKHKYRLEFDQFT	360
25	orf88a.pep	SMNVEDMSEGAEREKSLKSTLNDVRAVTQEGKKYTNIGPSIVYRIRDAAGQAVEYKNYML	420
	orf88-1	SMNVEDMSEGAEREKSLKSTLNDVRAVTQEGKKYTNIGPSIVYRIRDAAGQAVEYKNYML	420
30	orf88a.pep	PVLQEQDYFWITGTRSGLQQQYRWLRIP LDKQLKADTFMALREFLKDGEGRKRLVADATK	480
	orf88-1	PVLQEQDYFWITGTRSGLQQQYRWLRIP LDKQLKADTFMALREFLKDGEGRKRLVADATK	480
	orf88a.pep	GAPAEIREQFMLAAENTLNIFAQKGYLGLDEFITSNIPKEQQDKMQGYFYEMLYGVMNAA	540
35	orf88-1	GAPAEIREQFMLAAENTLNIFAQKGYLGLDEFITSNIPKEQQDKMQGYFYEMLYGVMNAA	540
	orf88a.pep	LDETIRRYGLPEWQQDEARNRFL LHSM DAYTGLTEYPAPMLLQLDGFSEVRSSGLQMTRS	600
40	orf88-1	LDETIRRYGLPEWQQDEARNRFL LHSM DAYTGLTEYPAPMLLQLDGFSEVRSSGLQMTRS	600
	orf88a.pep	PGALLVYLGSVLLVLTGTVLMFYVREKRAWVLFSDGKIRFAMSSARSERDLQKEFPKHVES	660
	orf88-1	PGALLVYLGSVLLVLTGTVLMFYVREKRAWVLFSDGKIRFAMSSARSERDLQKEFPKHVES	660
45	orf88a.pep	LQRLGKDLNHD	672
	orf88-1	LQRLGKDLNHD	672

### Homology with a predicted ORF from *N.gonorrhoeae*

50 ORF88 shows 93.8% identity over a 371aa overlap with a predicted ORF (ORF88.ng) from *N.*

### *gonorrhoeae*:

	orf88.pep	MVFLNADNGILVQDL PFEVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLERTIRVNH	60
	orf88ng	MVFLNADNGMLVQDL PFEVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLERTIRVNH	60
55	orf88.pep	PLTLHGITIYQASFADGGSDLT FKA WNLGDASREP VVLKATSIHQFPLEIGKHKYRLEFD	120
	orf88ng	PLTLHGITIYQASFADGGSDLT FKA WNLGDASREP VVLKATSIHQFPLEIGKHKYRLEFD	120
60	orf88.pep	QFTSMNVEDMSEGAEREKSLKSTLPDVRAVTQEGHKYTNXXXXXXYRIRDAPGQAVEYKN	180
	orf88ng	QFTSMNVEDMSEGAEREKSLKSTLNDVRAVTQEGKKYTNIGPSIVYRIRDAAGQAVEYKN	180
65	orf88.pep	YMLPVLQEQDYFWITGTRSLQQQYRWLRIP LDKQLKADTFMALREFLKDGEGRKRXVAD	240
	orf88ng	YMLPILQDKDYFWLTGTRSGLQQQYRWLRIP LDKQLKADTFMALREFLKDGEGRKRLVAD	240

5	orf88.pep	ATKGAPAEIREQFMLAAENTLNI FAQKGYLGLDEFITSNIPKEQQDKMQGYFYEMLYGVM	300
	orf88ng	ATKDAPAEIREQFMLAAENTLNI FAQKGYLGLDEFITSNIPKGQQDKMQGYFYEMLYGVM	300
10	orf88.pep	NAALDETXYTRYGLPEWQQDEARNRFLHSM DAYTGLTEYPAPMLLQLDGFSEVRSSGLQM	360
	orf88ng	NAALDETIRRYGLPEWQQDEARNRFLHSM DAYTGLTEYPAPMLLQLDGFSEVRSSGLQM	360
15	orf88.pep	TRSXGPLLVL	371
	orf88ng	TRSPGALLVYLGSVLLVLGTVFMFYVPPKKRAWVLF SNXKIRFAMSSARSERDLQKEFPKH	420

An ORF88ng nucleotide sequence <SEQ ID 333> was predicted to encode a protein having amino acid sequence <SEQ ID 334>:

15	1	MVFLNADNGM	LVQDLPFVEVK	LKKFHIDFYN	TGMPRDFASD	IEVTDKATGE
	51	KLERTIRVNH	PLTLHGITYI	QASFADGGS	LTFKAWNLRD	ASREPVVLKA
	101	TSIHQFPLEI	GKHKYRLEFD	QFTSMNVEDM	SEGAEREKSL	KSTLNDVRAV
	151	TQEGKKYTN	GPSIVYRIRD	AAGQAVEYKN	YMLPILQDKD	YFWLTGTRSG
	201	LQQQYRWLRI	PLDKQLKADT	FMALREFLKD	GEGRKRLVAD	ATKDAPAEIR
20	251	EQFMLAAENT	LNIFAQKGYL	GLDEFITSNI	PKGQQDKMQG	YFYEMLYGVM
	301	NAALDETIRR	YGLPEWQQDE	ARNRFLHSM	DAYTGLTEYP	APMLLQLDGF
	351	SEVRSSGLQM	TRSPGALLVY	LGSVLLVLGT	VFMFYVPPKR	AWVLF SNXKI
	401	RFAMSSARSE	RDLQKEFPKH	VESLQRLGKD	LNHD*	

Further work revealed the complete gonococcal DNA sequence <SEQ ID 335>:

25	1	ATGAGTAAAT	CCCGTATATC	TCCCACACTT	CTTCCCGTC	CGTGGTTCGC
	51	TTTTTTCAGC	TCCATGCGCT	TTGCGGTCGC	TTTGCTCAGT	CTGCTGGGTA
	101	TTGCATCGGT	TATCGGCACG	GTGTTACAGC	AAAACCAGCC	GCAGACGGAT
	151	TATTTGGTCA	AATTCGGACC	GTTTGGACT	CGGATTTTTC	ATTTTGGGG
	201	TTGTATGAT	GTCTATGCTT	CGGCATGGTT	TGTCGTTATC	ATGATGTTTC
30	251	TGGTGGTTC	TACCAAGTTG	TGTTAATCC	GTAACGTTCC	GCCGTTTGG
	301	CGCGAAATGA	AGTCTTTCG	GGAAAAGGTT	AAAGAAAAAT	CTCTGGCGGC
	351	GATGCGCCAT	TCTTCGCTGT	TGGATGTAAA	AATTGCCCCC	GAAGTTGCCA
	401	AACGTTATCT	AGGCGTCCG	GGTTTTCAGG	GAAAAACCGT	CAGCCGTGAG
	451	GACGGGTCCG	TTCTGATTGC	CGCCAAAAAA	GGCAcaatga	acaaATGGGG
35	501	CTATATCTTT	GCcgaagtag	ctTTGATTGT	CATTGCGCTG	GGCGGTTTGA
	551	TAGACAGTAA	CCTGCTGCTG	AAGCTGGGTA	TGCTGGCCGG	TCCGATTGTT
	601	CGGACAAATC	AGCGGTTTTC	TGCCAAGGAT	TTCAAGCCCG	AAAGTATTTT
	651	GGGTGCGTCC	AATCTCTCAT	TTAGGGGCAA	CGTCAATATT	TCCGAGGGGC
	701	AAAGTGCGGA	TGTGGTTTTTC	CTGAATGCCG	ACAACGGGAT	GTTGGTTTCA
40	751	GACTTGCTTT	TTGAAGTCAA	ACTGAAAAAA	TTCCATATCG	ATTTTATCAA
	801	TACGGGTATG	CCGCGCGATT	TTGCCAGCGA	TATTGAAGTA	ACGGACAAGG
	851	CAACCGGTGA	GAACTCGAG	CGCACCATCC	GCGTGAACCA	TCCTTTGACC
	901	TTGCACGGCA	TCACGATTTA	TCAGGCGAGT	TTTGCCGACG	GCGGTTCCGA
	951	TTTGACATTC	AAGCGGTGGA	ATTTGAGGGA	TGCTTCGCGC	GAACCTGTCT
45	1001	TGTTGAAGGC	AACCTCCATA	CACCAGTTTC	CGTTGGAAAT	CGGCAAAACAC
	1051	AAATATCGTC	TTGAGTTTCA	TCAGTTCACT	TCTATGAATG	TGGAGGACAT
	1101	GAGCGAGGGT	GCGGAACGGG	AAAAAAGCCT	GAAATCCACT	CTGAACGATG
	1151	TCCGCGCCGT	TACTCAGGAA	GGTAAAAAAT	ACACCAATAT	CGGCCCTTCC
	1201	ATCGTGTACC	GCATCCGTGA	TGcggCAGGG	CAGGCGGTCTG	AATATAAAAA
50	1251	CTATATGCTG	CCGATTTTGC	AGGACAAAGA	TTATTTTGG	CTGACCGGCA
	1301	CGCGCAGCGG	CTTCGAGCAG	CAATACCGCT	GGCTGCGTAT	CCCCTTGGAC
	1351	AAGCAGTTGA	AAGCGGACAC	CTTTATGGCA	TTGCGTGAGT	TTTTGAAAGA
	1401	TGGGGAAGGG	CGCAAACGTC	TGGTTGCCGA	CGCAACCAAA	GACGCACCTG
	1451	CCGAAATCCG	CGAACAATTC	ATGCTGGCTG	CGGAAAACAC	GCTGAATATC
55	1501	TTTGCACAAA	AAGGCTATTT	GGGATTGGAC	GAATTATTA	CGTCCAAATAT
	1551	CCCAGAAAGG	CAGCAGGATA	AGATGCAGGG	CTATTCTTAC	GAAATGCTTT
	1601	ACGGCGTGAT	GAACGCTGCT	TTGGATGAAA	CCATACGCCG	GTACGGCTTG
	1651	CCCGAATGGC	AGCAGGATGA	AGCGCGGAAC	CGTTTCCTGC	TGCACAGTAT
	1701	GATGCTTAT	ACGGGGCTGA	CGGAATATCC	CGCGCCTATG	CTGCTCCAGC
60	1751	TTGACGGGTT	TTCCGAGGTG	CGTTCCTCAG	GTTTGAGAT	GACCCGTTCG
	1801	CCGGGTGCGC	TTTTGGTCTA	Tctcggtctg	gtattgttgg	TTTTGGgtac
	1851	ggtatTttatg	tTTTATGTGC	GCGAAAAACG	GGCGTGGgta	tGTTTTTCag
	1901	aCGGCAAAAT	CCGTTTTGCT	ATGtCTTcgg	CCcgcagcga	ACGGGATTTG
	1951	cAGAgaggaaT	TTCCAAAACA	CGtcgAGAGC	CTGCAACggc	tcggcaaggga

2001 CttqaaTCAT GACTga

This corresponds to the amino acid sequence <SEQ ID 336; ORF88ng-1>:

	1	MSKSRISPTL	LSRPWFAFFS	SMRFAVALLS	LLGIASVIGT	VLQONQPQTD
5	51	YLVKFGPFWT	RIFDFLGLYD	VYASAWFVVI	MMFLVVSSTSL	CLIRNVPPFW
	101	REMKSFREKV	KEKSLAAMRH	SSLDDVKIAP	EVAKRYLEVR	GFQKGTVSRE
	151	DGSVLIAAKK	GTMNKGGYIF	AQVALIVICL	GGILIDNLLR	LKGLMAGRIV
	201	PDNQAVYAKD	FKPESILGAS	NLSFRGNVNI	SEQQSADVVF	LNADNGMLVU
	251	DLPEFVKLKK	FHIDFYNTGM	PRDFASDIEV	TOKATGKESI	RTIRVGNHPLT
10	301	LHGITIYQAS	FADGGSDLTF	KAWNLRDASR	EPVVLKATLE	HQFPLEIGKH
	351	KYRLEFDQST	SMNVEDMSEG	AEREKSLKST	LNDVRAVTQE	GKKYTNIGPS
	401	IVYRIRDAAG	QAVEYKNYML	PILQDKDIYF	LTGTSRGLQG	QYRWLRILPLD
	451	QQLKADTFMA	LREFLKDGEG	RKRLVADATK	DPAEAIREQF	MLAAENTLNI
	501	FAQKGYLGLD	EFITSNIPKG	QQDKMQGYFY	EMLYGVMNAA	LDETIRRYGL
15	551	PEWQQDEARN	RFLLSMDAY	TGLTEYPAPM	LLQLDGGFSEV	RSSGLQMTRS
	601	PGALLVYLGS	VLLVLGTVMF	FYVREKRAWV	LFSDGKIRFA	MSSARSERDL
	651	OKEFPKHVES	LQRLGKDLNH	D*		

ORF88ng-1 and ORF88-1 show 97.0% identity in 671 aa overlap:

	orf88-1.pep	MSKSRSPPLLSRPWFAFFSSMRFAVALLSLLGIASVIGTVLQQNQPQTDYLVKFGSFWA 	60
20	orf88ng-1	MSKSRI SPTLLSRPWFAFFSSMRFAVALLSLLGIASVIGTVLQQNQPQTDYLVKFGPFWT    :	60
	orf88-1.pep	QIFGFLGLYDVYASAWFVVIMMFLVVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH :	120
25	orf88ng-1	RIFDFLGLYDVYASAWFVVIMMFLVVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH    :	120
	orf88-1.pep	SSLDDVKIAPEVAKRYLEVGQFQGKTINREDGSVLIAAKKGTMNKGWYIFAHVALIVICL    :::	180
	orf88ng-1	SSLDDVKIAPEVAKRYLEVRGFQGTVSREDGSVLIAAKKGTMNKGWYIFAQVALIVICL    :::	180
30	orf88-1.pep	GGLIDSNLLLKLGMALTGRIVPDNQAVYAKDFKPESILGASNLSFRGNVNISEGQSADVVF    :::	240
	orf88ng-1	GGLIDSNLLLKLGMLAGRIVPDNQAVYAKDFKPESILGASNLSFRGNVNISEGQSADVVF    :::	240
35	orf88-1.pep	LNADNGILVQDLPFEVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLETRIRNVHPLT    :::	300
	orf88ng-1	LNADNGMLVQDLPFEVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLETRIRNVHPLT    :::	300
	orf88-1.pep	LHGITIYQASFADGGSDLTTFKAWNLGDAAREPVVLKATSIHQFPLEIGKHKYRLEFDQFT    :::	360
40	orf88ng-1	LHGITIYQASFADGGSDLTTFKAWNLRDAAREPVVLKATSIHQFPLEIGKHKYRLEFDQFT    :::	360
	orf88-1.pep	SMNVEDMSEGAEREKSLKSTLNDVRAVTQEGKKYTNI GPSIVYRIRDAAGQAVEYKNMYL    :::	420
45	orf88ng-1	SMNVEDMSEGAEREKSLKSTLNDVRAVTQEGKKYTNI GPSIVYRIRDAAGQAVEYKNMYL    :::	420
	orf88-1.pep	PVLQE QDYFWITGTRSGLQQQYRWLRIPLDKQLKADTFMALREFLKDGEGRKRLVADATK  :::	480
	orf88ng-1	PILQDKDYFWLTGTRSGLQQQYRWLRIPLDKQLKADTFMALREFLKDGEGRKRLVADATK  :::	480
50	orf88-1.pep	GAPAEIREQFM LAAENTLNIFAQKGYLGLDEFITSNIPKEQQDKMQGYFYEMLYGVMNAA    :::	540
	orf88ng-1	DAPAEIREQFM LAAENTLNIFAQKGYLGLDEFITSNIPKGQQDKMQGYFYEMLYGVMNAA    :::	540
55	orf88-1.pep	LDETIRRYGLPEWQQDEARNRFL LHSMDAYTGLTEYPAPMLLQLDGFSEVRSSGLQMTRS    :::	600
	orf88ng-1	LDETIRRYGLPEWQQDEARNRFL LHSMDAYTGLTEYPAPMLLQLDGFSEVRSSGLQMTRS    :::	600
	orf88-1.pep	PGALLVYLGSVLLVLGTVMFYVREKRAWVLFSDGKIRFAMSSARSERDLQKEFKPHVES    :::	660
60	orf88ng-1	PGALLVYLGSVLLVLGTVMFYVREKRAWVLFSDGKIRFAMSSARSERDLQKEFKPHVES    :::	660
	orf88-1.pep	LQRLGKDLNHD      671 	
	orf88ng-1	LQRLGKDLNHD      671 	

Furthermore, ORG88ng-1 shows homology with a hypothetical protein from *Aquifex aeolicus*:

gi|2984296 (AE000771) hypothetical protein [Aquifex aeolicus] Length = 537  
 Score = 94.4 bits (231), Expect = 2e-18  
 Identities = 91/334 (27%), Positives = 159/334 (47%), Gaps = 59/334 (17%)

5 Query: 16 FAFFSSMRFAVALLSLGLIASVIG-TVLQQNQPTDYLKFGPFWTRIFDFLGLYDVYAS 74  
 + F +S++ A+ ++ +LGI S++G T ++QNQ YL +FG L L DV+ S  
 Sbjct: 80 YDFLASLKLAI FIMLVLGILSMLGSTYIKQNQSFEWYLDQFGYDVGIWIWKLWLNDFHS 139

10 Query: 75 AWFVVMFLVSTSLCLIRNVPFWMKSFREKVKEKSLAAMRHSSLLDVKIAPEVAK 134  
 ++++ ++ L V+ C I+ +P W++ S +E++ + A +H + VKI P+ K  
 Sbjct: 140 WYYILFIVLLAVNLIFCSIKRLPRVWKQAFS-KERILKLDEHAEKHLKPITVKI-PDKK 197

15 Query: 135 --RYLEVRGFGKTVSREDGSVLIAAKKGTMNKNGYIFAQVALIVICLGLIDSNLLKL 192  
 ++L +GF+ V E + + A+KG ++ G +AL+VI G LID  
 Sbjct: 198 VLKFLKKGFK-VFVEEENKLYVFAEKGRFSRLGVYITHIALLVIMAGALID----- 249

20 Query: 193 GMLAGRIVPDNQAVYAKDFKPEISILGASNLSFRGNVNISEGQSADVFLNADNGMLVQDL 252  
 +I+G RG++ ++EG + DV+ + A+ L  
 Sbjct: 250 -----AIVGV-----RGLIVAEGDTNDVMLVGAE--QKPYKL 280

25 Query: 253 PFEVKLKKFHIDFY---NTGMPRDFA-----SDIEVTDKATGEKLER--TIRVNHPLT 300  
 PF V L F I Y N + + FA SDIE+ + G K+E T++VN P  
 Sbjct: 281 PFAVHLIDFRIKTYAEENPNVDRFAQAVSSYESDIEIIN---GGKVEAKGTVKVNEPFD 337

Query: 301 LHGITIYQASFA--DGGSDLTFFKAWNLRDASREP 332  
 ++QA++ DG S + + + A +P  
 Sbjct: 338 FGRYRLFQATYGILDGTSGMGVIVVDRKKAHEDP 371

Based on this analysis, including the putative transmembrane domain in the gonococcal protein,  
 30 it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could  
 be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 40

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID  
 337>:

35 1 ATGATGAGTA ATAmAATGGm ACAAAAAGGG TTTACATTGA TTGmGmTGAT  
 51 GATAGTCGTC GCGATACTCG GCATTATCAG CGTCATTGCC ATACCTTCTT  
 101 ATCmAAGTTA TATTGAAAAA GGCTATCAGT CCCAGCTTTA TACGGAGATG  
 151 GyCGGTATCA ACAATATTTC CAAACAGTTT ATTTTGAAAA ATCCCCTGGA  
 201 CGATAATCAG ACCATCGAGA ACAAAGTGA AATATTTGTC TCAGGCTATA  
 40 251 AGATGAATCC GAAAATTGCC AAAAAaTATA GTGTTTCGGT AAAGTTTGTC  
 301 GATAAGGAAA AATCAAGGCG ATACAGGTTG GTCGGCGTTC CGAAGGCGGG  
 351 GACGGGTAT ACTTTGTCGG TATGGATGAA CAGCGTGGGC GACGGATACA  
 401 AATGCCGTGA TGCCGCTTCT GCCCAAGCCC ATTTGGAGAC CTTGTCCTCA  
 451 GATGTCGGCT GTGAAGCCTT CTCTAATCGT AAAAAATAA

45 This corresponds to the amino acid sequence <SEQ ID 338; ORF89>:

1 MMSNXMXQKG FTLIXXMIVV AILGIISVIA IPSYXSYIEK GYQSQLYTEM  
 51 XGINNISKQF ILKNPLDDNQ TIENKLEIFV SGYKMNPKIA KKYSVSVKVF  
 101 DKEKSRAAYRL VGVPKAGTGY TLSVWMNSVG DGYKCRDAAS AQAHLETLS  
 151 DVGCEAFSNR KK\*

50 Further work revealed the complete nucleotide sequence <SEQ ID 339>:

1 ATGATGAGTA ATAAAATGGA ACAAAAAGGG TTTACATTGA TTGAGATGAT  
 51 GATAGTCGTC GCGATACTCG GCATTATCAG CGTCATTGCC ATACCTTCTT  
 101 ATCAAAGTTA TATTGAAAAA GGCTATCAGT CCCAGCTTTA TACGGAGATG  
 151 GTCGGTATCA ACAATATTTC CAAACAGTTT ATTTTGAAAA ATCCCCTGGA

201 CGATAATCAG ACCATCGAGA ACAAATGGA AATATTTGTC TCAGGCTATA  
 251 AGATGAATCC GAAAATTGCC AAAAAATATA GTGTTTCGGT AAAGTTTGTC  
 301 GATAAGGAAA AATCAAGGGC ATACAGGTTG GTCGGCGTTC CGAAGGCGGG  
 351 GACGGGTTAT ACTTTGTTCG TATGGATGAA CAGCGTGGGC GACGGATACA  
 401 AATGCCGTGA TGCCGCTTCT GCCCAAGCCC ATTTGGAGAC CTTGTCCTCA  
 451 GATGTCGGCT GTGAAGCCTT CTCTAATCGT AAAAAATAA

This corresponds to the amino acid sequence <SEQ ID 340; ORF89-1>:

1 MMSNKMEQKG FTLEMMIVV AILGIISVIA IPSYQSYIEK GYQSOLYTEM  
 51 VGINNISKQF ILKNPLDDNQ TIENKLEIFV SGYKMNPKIA KKYSVSVKVF  
 101 DKEKSRAYRL VGVPKAGTGY TSVWMNSVG DGYKCRDAAS AQAHLETLS  
 151 DVGCEAFSNR KK\*

Computer analysis of this amino acid sequence gave the following results:

Homology with Pile of *N. gonorrhoeae* (accession number Z69260).

ORF89 and Pile protein show 30% aa identity in 120a overlap:

15 orf89 8 QKGFTLIXXMIVVAILGIISVIAIPSYKSYIEKGYQSOLYTEMXGINNISKQFILKNPL- 66  
 QKGFTLI MIV+AI+GI++ +A+P+Y Y + S+ G + ++ L + +  
 Pile 5 QKGFTLIELMIVIAIVGILAAVALPAYQDYTARAQVSEAILLAEGQKSAVTEYYLNHGIW 64  
 20 orf89 67 -DDNQTIENKLEIFVSGYKMNPKIAKKYSVSVKVFVDKEKSRAYRLVGVPKAGTGYTLVSW 125  
 DN + +G + KI KY SV + GV K G LS+W  
 PILE 65 PKDNTS-----AGVASSDKIKGYQSVTVAKGVVTAEMASTGVNKEIQGKKLSLW 115

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF89 shows 83.3% identity over a 162aa overlap with an ORF (ORF89a) from strain A of *N. meningitidis*:

25 orf89.pep 10 20 30 40 50 60  
 MMSNXMXQKGF TLIXXMIVVAILGIISVIAIPSYKSYIEKGYQSOLYTEMXGINNISKQF  
 orf89a 10 20 30 40 50 60  
 MMSNKMEQKGFTLIXXXXXAIXXXXSVIXXXXSYIEKGYQSOLYTEMVGINNISKQX  
 30 orf89.pep 70 80 90 100 110 120  
 ILKNPLDDNQTIENKLEIFVSGYKMNPKIAKKYSVSVKVFVDKEKSRAYRLVGVPKAGTGY  
 orf89a 70 80 90 100 110 120  
 ILKNPLDDNQTIKSKLEIFVSGYKMNPKIAEKYNVSVHFNVEEKPRAYSLVGVPKGTGY  
 35 orf89.pep 130 140 150 160  
 TSVWMNSVGDGYKCRDAASAQALETLSDDVGCEAFSNRKKX  
 orf89a 130 140 150 160  
 TSVWMNSVGDGYKCRDAASARALETLSDDVGCEAFSNRKKX

The complete length ORF89a nucleotide sequence <SEQ ID 341> is:

1 ATGATGAGTA ATAAATGGA ACAAAGGG TTTACATTGA TTGNGANGNT  
 45 51 NATNGNCNTC GCGATACNCN GCNTTANCAG CGTCATTNCN ATNNNTNCNT  
 101 ATCNNAGTTA TATTGAAAAA GGCTATCAGT CCCAGCTTTA TACGGAGATG  
 151 GTCGGTATCA ACAATATTTC CAAACAGTNT ATTTTGAAAA ATCCCTGGGA  
 201 CGATAATCAG ACCATCAAGA GCAAATGGA AATATTTGTC TCAGGCTATA  
 251 AGATGAATCC GAAAATTGCC AAAAAATATA ATGTTTCGGT GCATTTTGTC  
 50 301 AATGAGGAAA AACNAGGGC ATACAGCTTG GTCGGCGTTC CAAAGACGGG  
 351 GACGGGTTAT ACTTTGTTCG TATGGATGAA CAGCGTGGGC GACGGATACA  
 401 AATGCCGTGA TGCCGCTTCT GCCCGAGCCC ATTTGGAGAC CTTGTCCTCA  
 451 GATGTCGGCT GTGAAGCCTT CTCTAATCGT AAAAAATAG

This encodes a protein having amino acid sequence <SEQ ID 342>:

55 1 MMSNKMEQKG FTLIXXXXXX AIXXXXSVIX XXXYXSYIEK GYQSOLYTEM

51 VGINNISKQX ILKNPLDDNQ TIKSKLEIFV SGYKMNPKIA EKYNVSVHFV  
 101 NEEKPRAYSL VGVPKTGTGY TLSVWMNSVG DGYKCRDAAS ARAHLETLS  
 151 DVGCEAFSNR KK\*

ORF89a and ORF89-1 show 83.3% identity in 162 aa overlap:

5		10	20	30	40	50	60
	orf89a.pep	MMSNKMEQKGFTLIXXXXXXAIXXXXSVIXXXYSYIEKGYQSOLYTEMVGINNISKQX					
	orf89-1	MMSNKMEQKGFTLIEMMIVVAILGIISVIAIPSYQSYIEKGYQSOLYTEMVGINNISKQF					
10		70	80	90	100	110	120
	orf89a.pep	ILKNPLDDNQTIKSKLEIFVSGYKMNPKIAEKYNVSVHFVNEEKPRAYSLVGVPKTGTGY					
	orf89-1	ILKNPLDDNQTIENKLEIFVSGYKMNPKIAKKYSVSVKFDKEKSRAYRLVGVPKAGTGY					
15		130	140	150	160		
	orf89a.pep	TLSVWMNSVG DGYKCRDAAS ARAHLETLS DVGCEAFSNRKKX					
20		130	140	150	160		
	orf89-1	TLSVWMNSVG DGYKCRDAAS ARAHLETLS DVGCEAFSNRKKX					

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF89 shows 84.6% identity over a 162aa overlap with a predicted ORF (ORF89.ng) from *N. gonorrhoeae*:

25	orf89	MMSNXMXQKGFTLIXXMIVVAILGIISVIAIPSYXSYIEKGYQSOLYTEMXGINNISKQF	60
	orf89ng	MMSNKMEQKGFTLIEMMIVVTILGIISVIAIPSYQSYIEKGYQSOLYTEMVGINNVLKQF	60
30	orf89	ILKNPLDDNQTIENKLEIFVSGYKMNPKIAKKYSVSVKFDKEKSRAYRLVGVPKAGTGY	120
	orf89ng	ILKNPQDDNDTLKSKLIFVSGYKMNPKIAKKYSVSVRFVDAEKPRAYRLVGVPNAGTGY	120
35	orf89	TLSVWMNSVG DGYKCRDAAS ARAHLETLS DVGCEAFSNRKK	162
	orf89ng	TLSVWMNSVG DGYKCRDATSAQAYSDTLSADSGCEAFSNRKK	162

The complete length ORF89ng nucleotide sequence <SEQ ID 343> is:

1	atGATGAGCA ATAAATGGA AAAAAAGGG TTTACATTGA TTGAGATGAT
51	GATAGTTGTC ACGATACTCG GCATCATCAG CGTCATTGCC ATACCTTCTT
101	ATCAGAGTTA TATTGAAAAA GGCTATCAGT CCCAGCTTTA TACGGAGATG
151	GTCGGTATCA ACAATGTTCT CAAACAGTTT ATTTTGAAAA ATCCCCAGGA
201	CGATAATGAT ACCCTCAAGA GCAAACTGAA AATATTTGTC TCAGGCTATA
251	AGATGAATCC GAAAAattgCC AAAAAATATA GTGTTTCGGt aaggtttGTC
301	gatGCGGAAA AACCAAGGGC ATACAGGTTG GTCGGCGTTC CGAACGCGGG
351	GACGGGTTAT ACTTTGTCGG TATGGATGAA CAGCGTGGGC GACGGATACA
401	AATGCCGTGA TGCCACTTCT GCCCAGGCCT ATTCGGACAC CTTGTCCGCA
451	GATAGCGGCT GTGAAGCTTT CTCTAATCGT AAAAAATAG

This encodes a protein having amino acid sequence <SEQ ID 344>:

1	MMSNKMEQKG FTLIEMMIVV TILGIISVIA IPSYQSYIEK GYQSOLYTEM
51	VGINNVLKQF ILKNPQDDND TLKSKLKIFV SGYKMNPKIA KKYSVSVRFV
101	DAEKPRAYRL VGVPNAGTGY TLSVWMNSVG DGYKCRDATS AQAYSDTLSA
151	DSGCEAFSNR KK*

This gonococcal protein has a putative leader peptide (underlined) and N-terminal methylation site (NMePhe or type-4 pili, double-underlined). In addition, ORF89ng and ORF89-1 show 88.3%

identity in 162 aa overlap:



		10	20	30	40	50	60
	orf89-1.pep	MMSNKMEQKGF	TLIEMMIVVAIL	GIISVIAIPSY	QSYIEKGYQ	SQLYTEMV	GINNISKQF
5	orf89ng	MMSNKMEQKGF	TLIEMMIVVIL	GIISVIAIPSY	QSYIEKGYQ	SQLYTEMV	GINNVLKQF
		10	20	30	40	50	60
		70	80	90	100	110	120
	orf89-1.pep	ILKNPLDDNQT	IENKLEIFV	SGYKMNPKI	AKKYSVSV	KFVDKEKS	RAYRLVGV
10	orf89ng	ILKNPQDDND	TLKSKLKI	FVSGYKMN	PKIAKKYS	SVSVRFV	DAEKPRAY
		70	80	90	100	110	120
		130	140	150	160		
	orf89-1.pep	TLSVWMNSV	GDGYKCRD	AASQAHL	ETLSSDV	GCEAFS	NRKKX
15	orf89ng	TLSVWMNSV	GDGYKCRD	ATSAQAY	SDTLSAD	SGCEAF	SNRKKX
		130	140	150	160		

Based on this analysis, including the gonococcal motifs and the homology with the known Pile protein, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF89-1 (13.6kDa) was cloned in the pGex vector and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 11A shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein was used to immunise mice, whose sera gave a positive result in the ELISA test., confirming that ORF89-1 is a surface-exposed protein, and that it is a useful immunogen.

#### Example 41

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 345>:

```

1  ATGAAAAAAT CCTCCCTCAT CAGCGCATTG GGCATCGGTA TTTTGAGCAT
51  CGGCATGGCA TTTGCCGCCC CTGCCGACGC GGTAAAGCCAA ATCCGTCAAA
101 ACGCCACTCA AGTATTGAGC ATCTTAAAAA ACGGCGATGC CAACACCGCT
151 CGCCAAAAAG CCGAAGCCTA TGCATTCCC TATTTCGATT TCCAACGTAT
201 GACCGCATTG GCGGTCGGCA ACCCTTGGsG CACCG.GTCC GACG.GCAAA
251 AACAAGCGTT GGCCn.AGAA TTCAACCC...
```

This corresponds to the amino acid sequence <SEQ ID 346; ORF91>:

```

1  MKKSSLISAL GIGILSIGMA FAAPADAVSQ IRQNATQVLS ILKNGDANTA
51  RQKAEAYAIP YFDFQRM TAL AVGNPWXTXS DXQKQALAXE FQP...
```

Further work revealed the complete nucleotide sequence <SEQ ID 347>:

```

1  ATGAAAAAAT CCTCCCTCAT CAGCGCATTG GGCATCGGTA TTTTGAGCAT
51  CGGCATGGCA TTTGCCGCCC CTGCCGACGC GGTAAAGCCAA ATCCGTCAAA
101 ACGCCACTCA AGTATTGAGC ATCTTAAAAA ACGGCGATGC CAACACCGCT
151 CGCCAAAAAG CCGAAGCCTA TGCATTCCC TATTTCGATT TCCAACGTAT
201 GACCGCATTG GCGGTCGGCA ACCCTTGGCG CACCGCGTCC GACGCGCAAA
251 AACAAGCGTT GGCCAAAGAA TTTCAAACCC TGCTGATCCG CACCTATTCC
301 GGCACGATGC TGAAATTAAA AAACGCCAAC GTCAACGTCA AAGACAATCC
351 CATCGTCAAT AAAGGCGGCA AAGAAATCAT CGTCCGCGCC GAAGTCGGCG
401 TACCCGGGCA AAAACCCGTC AACATGGACT TCACCACCTA CCAAGCGGC
451 GGTAAATACC GTACCTACAA CGTCGCCATC GAAGGCGCGA GCCTGGTTAC
501 CGTGTACCGC AACCAATTCG GCGAAATTAT CAAAGCGAAA GCGGTGGACG
551 GACTGATTGC CGAGTTGAAA GCCAAAAACG GCGGCAAATA A
```

This corresponds to the amino acid sequence <SEQ ID 348; ORF91-1>:

```

      1 MKKSSLISAL GIGILSIGMA FAAPADAVSQ IRQNATQVLS ILKNGDANTA
     51 RQKAEAYAIP YFDFQRM TAL AVGNPWRTAS DAQKQALAKE FQTLIRTYYS
    101 GTMLKLKNAN VNVKDNPIVN KGGKEIIVRA EVGVPGQKPV NMDFTTYQSG
    151 GKYRTYNVAI EGASLVTYR NQFGEI IKAK GVDGLIAELK AKNGGK*
  
```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF91 shows 92.4% identity over a 92aa overlap with an ORF (ORF91a) from strain A of *N.*

*meningitidis*:

```

10      orf91.pep      10      20      30      40      50      60
      MKKSSLISALGIGILSIGMAFAAPADAVSQIRQNATQVLSILKNGDANTARQKAEAYAIP
      orf91a          10      20      30      40      50      60
      MKKSSFISALGIGILSIGMAFAAPADAVNQIRQNATQVLSILKSGDANTARQKAEAYAIP

15      orf91.pep      70      80      90
      YFDFQRM TAL AVGNPWXTXSDXQKQALAXEFQP
      orf91a          70      80      90      100      110      120
      YFDFQRM TAL AVGNPWRTASDAQKQALAKEFQTLIRTYSGTMLKLKNANVNVKDNPIVN

20      orf91a          70      80      90      100      110      120
      KGGKEIIVRAEVLPGQKPVNMDFTTYQSGGKYRTYNVAIEGASLVTYRNVNMFGEI IKAK
      130      140      150      160      170      180
  
```

The complete length ORF91a nucleotide sequence <SEQ ID 349> is:

```

25      1 ATGAAAAAAT CCTCCTTCAT CAGCGCATTG GGCATCGGTA TTTGAGCAT
     51 CGGCATGGCA TTTGCCGCCCT CTGCCGACGC GGTAAACCAA ATCCGTCAA
    101 ACGCCACTCA AGTATTGAGC ATCTTAAAAA GCGGTGATGC CAACACCGCC
    151 CGCCAAAAAG CCGAAGCCTA TGCATTCCG TATTTCGATT TCCAACGTAT
    201 GACCGCATTG GCGGTCGGCA ACCCTGGCG CACCGCGTCC GACGCGCAA
    30 251 AACAAAGCGTT GGCCAAAGAA TTTCAAACCC TGCTGATCCG CACCTATTCC
     301 GGCACGATGC TGAAATTAAA AAACGCCAAC GTCAACGTCA AAGACAATCC
     351 CATCGTCAAT AAAGGCGGCA AAGAAATCAT CGTCCGCGCC GAAGTCGGCG
     401 TACCCGGGCA AAAACCCGTC AACATGGACT TCACCACCTA CCAAAGCGGC
     451 GGTAAATACC GTACCTACAA CGTCGCCATC GAAGGCGCGA GCCTGGTTAC
    35 501 CGTGTACCGC AACCAATTCG GCGAAATTAT CAAAGCGAAA GCGGTGGACG
     551 GACTGATTGC CGAGTTGAAG GCTAAAAACG GCAGCAAGTA A
  
```

This encodes a protein having amino acid sequence <SEQ ID 350>:

```

40      1 MKKSSFISAL GIGILSIGMA FAAPADAVNQ IRQNATQVLS ILKSGDANTA
     51 RQKAEAYAIP YFDFQRM TAL AVGNPWRTAS DAQKQALAKE FQTLIRTYYS
    101 GTMLKLKNAN VNVKDNPIVN KGGKEIIVRA EVGVPGQKPV NMDFTTYQSG
    151 GKYRTYNVAI EGASLVTYR NQFGEI IKAK GVDGLIAELK AKNGSK*
  
```

ORF91a and ORF91-1 show 98.0% identity in 196 aa overlap:

```

45      orf91a.pep      10      20      30      40      50      60
      MKKSSFISALGIGILSIGMAFAAPADAVNQIRQNATQVLSILKSGDANTARQKAEAYAIP
      orf91-1          10      20      30      40      50      60
      MKKSSLISALGIGILSIGMAFAAPADAVSQIRQNATQVLSILKNGDANTARQKAEAYAIP

50      orf91a.pep      70      80      90      100      110      120
      YFDFQRM TAL AVGNPWRTASDAQKQALAKEFQTLIRTYSGTMLKLKNANVNVKDNPIVN
      orf91-1          70      80      90      100      110      120
      YFDFQRM TAL AVGNPWRTASDAQKQALAKEFQTLIRTYSGTMLKLKNANVNVKDNPIVN

55      orf91a.pep      130      140      150      160      170      180
      KGGKEIIVRAEVLPGQKPVNMDFTTYQSGGKYRTYNVAIEGASLVTYRNVNMFGEI IKAK
  
```

orf91-1 KGGKEIIVRAE VGVPGQKPVNMDFTTYQSGGKYRTYNVAIEGASLVTYVRNQFGEI I KAK  
130 140 150 160 170 180

5

orf91a.pep GVDGLIAELKAKNGSKX  
190  
orf91-1 GVDGLIAELKAKNGGKX  
190

10

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF91 shows 84.8% identity over a 92aa overlap with a predicted ORF (ORF91.ng) from *N. gonorrhoeae*:

15 orf91.pep MKKSSLISALGIGILSIGMAFAAPADAVSQIRQATQVLSILKNGDANTARQKAEAYAIIP 60  
orf91ng VKKSSFISALGIGILSIGMAFASPADAVGQIRQATQVLTILKSGDAASARPKAEAYAVP 60  
orf91.pep YFDFQRM TALAVGNPWXTXSDXQKQALAXEFQP 93  
20 orf91ng YFDFQRM TALAVGNPWRTASDAQKQALAKEFQTL LIRTYSGTMLKFKNATVNVKDNPIVN 120

The complete length ORF91ng nucleotide sequence <SEQ ID 351> is predicted to encode a protein having amino acid sequence <SEQ ID 352>:

25 1 VKKSSFISAL GIGILSIGMA FASPADAVGQ IRQATQVLT ILKSGDAASA  
51 RPKAEAYAVP YFDFQRM TAL AVGNPWRTAS DAQKQALAKE FQTL LIRTYSG  
101 GTMLKFKNAT VNVKDNPIVN KGGKEIVVRA EVGIPGQKPV NMDFTTYQSG  
151 GKYRTYNVAI EGTS LVTYVR NQFGEI I KAK GIDGLIAELK AKNGGK\*

Further work revealed the complete nucleotide sequence <SEQ ID 353>:

30 1 ATGAAAAAAT CCTCCTTCAT CAGCGCATTG GGCATCGGTA TTTTGAGCAT  
51 CGGCATGGCA TTTGCCTCCC CGGCCGACGC AGTGGGACAA ATCCGCCAAA  
101 ACGCCACACA GGTTTTGACC ATCCTCAAAA GCGGCGACGC GGCTTCTGSA  
151 CGCCCAAAAG CCGAAGCCTA TGCGGTTCCT TATTTCGATT TCCAACGTAT  
201 GACCGCATTG GCGGTCGGCA ACCCTTGGCG TACCGCGTCC GACGCGCAAA  
251 AACAAAGCGT GGCCAAAGAA TTTCAAACCC TGCTGATCCG CACCTATTCC  
301 GGCACGATGC TGA AATTCAA AAACGCGACC GTCAACGTCA AAGACAATCC  
351 CATCGTCAAT AAGGGCGGCA AGGAAATCGT CGTCCGTGCC GAAGTCGGCA  
401 TCCCCGGTCA GAAGCCCGTC AATATGGACT TTACCACCTA CCAAAGCGGC  
451 GGCA AATACC GTACCTACAA CGTCGCCATC GAAGGCACGA GCCTGGTTAC  
501 CGTG TACCGC AACCAATTCG GCGAAATCAT CAAAGCCAAA GGCATCGACG  
551 GGCTGATTGC CGAGTTGAAA GCCAAAACG GCGGCA AATA A

40 This corresponds to the amino acid sequence <SEQ ID 354; ORF91ng-1>:

1 MKKSSFISAL GIGILSIGMA FASPADAVGQ IRQATQVLT ILKSGDAASA  
51 RPKAEAYAVP YFDFQRM TAL AVGNPWRTAS DAQKQALAKE FQTL LIRTYSG  
101 GTMLKFKNAT VNVKDNPIVN KGGKEIVVRA EVGIPGQKPV NMDFTTYQSG  
151 GKYRTYNVAI EGTS LVTYVR NQFGEI I KAK GIDGLIAELK AKNGGK\*

45 ORF91ng-1 and ORF91-1 show 92.3% identity in 196 aa overlap:

50 orf91-1.pep MKKSSLISALGIGILSIGMAFAAPADAVSQIRQATQVLSILKNGDANTARQKAEAYAIIP  
orf91ng-1 MKKSSFISALGIGILSIGMAFASPADAVGQIRQATQVLTILKSGDAASARPKAEAYAVP  
orf91-1.pep YFDFQRM TALAVGNPWRTASDAQKQALAKEFQTL LIRTYSGTMLKFKNATVNVKDNPIVN  
55 orf91ng-1 YFDFQRM TALAVGNPWRTASDAQKQALAKEFQTL LIRTYSGTMLKFKNATVNVKDNPIVN

```

                    130      140      150      160      170      180
orf91-1.pep  KGGKEIIVRAEVGVPGQKPVNMDFTTYQSGGKYRTYNVAIEGASLVTVYRNQFGEI IKAK
5  orf91ng-1  KGGKEIIVRAEVGIPGQKPVNMDFTTYQSGGKYRTYNVAIEGTSLVTVYRNQFGEI IKAK
                    130      140      150      160      170      180

                    190
orf91-1.pep  GVDGLIAELKAKNGGKX
10 orf91ng-1  GIDGLIAELKAKNGGKX
                    190

```

In addition, ORF91ng-1 shows homology to a hypothetical *E.coli* protein:

```

15  sp|P45390|YRBC_ECOLI_HYPOTHETICAL_24.0_KD_PROTEIN_IN_MURA-RPON_INTERGENIC
    REGION_PRECURSOR_(F211)>gi|606130|(U18997)ORF_f211[Escherichia coli]
    >gi|1789583|(AE000399)hypothetical_24.0_kD_protein_in_murZ-rpoN_intergenic
    region[Escherichia coli]Length = 211

```

```

20  Score = 70.6 bits (170), Expect = 6e-12
    Identities = 42/137 (30%), Positives = 76/137 (54%), Gaps = 6/137 (4%)

```

```

Query: 59 VPFDFQRM TALAVGNPWRTASDAQKQALAKEFQTLIRTYSGMTLKFKNATNVNVDNPI 118
        +PY + AL +G +++A+ AQ++A F+ L + Y + + T + P
Sbjct: 65 LPYVQVKYAGALVLGQYYKSATPAQREAYFAAFREYLKQAYGQALAMYHGQTYQIA--PE 122

```

```

25  Query: 119 VNKGKKEIV-VRAEVGIP-GQKPVNMDFTTYQSG--GKYRTYNVAIEGTSLVTVYRNQFG 174
        G K IV +R + P G+ PV +DF ++ G ++ Y++ EG S++T +N++G
Sbjct: 123 QPLGDKTIVPIRVTIIDPNGRPPVRLDFQWRKNSQTGNWQAYDMIAEGVSMITTKQNEWG 182

```

```

30  Query: 175 EIIKAKGIDGLIAELKA 191
        +++ KGIDGL A+LK+
Sbjct: 183 TLLRTKGIDGLTAQLKS 199

```

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 42

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 355>:

```

40  1 ATGAAACACA TACTCCCCCT GATTGCCGCA TCCGCACTCT GCATTTC AAC
    51 CGCTTCGGCA CATCCTGCCA GCGAACCGTC CACTCAAAAC GAAACCGCTA
    101 TGATCACGCA TACCCTCATC TCAAAATACA GTTTTGnnn nnnnnnnnnn
    151 nnnnnnnnnn nnGCCATAAA AAGCAAAGGG ATGGACATTT TTGCCGTCAT
    201 CGACCATCAG GAAGCCGCAC GCCGAAACGG CTTAACGATG CAGCCGGCAA
    251 AAGTCATCGT CTTCGGCACG CCCAAAGCCG GCACGCCGCT GATGGTCAAA
    301 GACCCCGCCT TCGCCCTGCA ACTGCCCTTA CGCGTCCTCG TTACCGAAAC
45  351 GGACGGCAAA GTACGCGCCG CCTATACCGA TACGCGCGCC CTCATCGCCG
    401 GCAGCCGCAT CGGTTTCGAC GAAGTGGCAA ACACTTTGGC AAACGCCGAA
    451 AAAGTGATAC AAAAAACCGT AGGCGAATAA

```

This corresponds to the amino acid sequence <SEQ ID 356; ORF97>:

```

50  1 MKHILPLIAA SALCISTASA HPASEPSTQN ETAMITHTLI SKYSFGXXXX
    51 XXXXAIKSKG MDIFAVIDHQ EAARRNGLTM QPAKIVVFGT PKAGTPLMVK
    101 DPAFALQLPL RVLVTETDGK VRAAYTDTRA LIAGSRIGFD EVANTLANAE
    151 KLIQKTVGE*

```

Further work revealed the complete nucleotide sequence <SEQ ID 357>:

```

55  1 ATGAAACACA TACTCCCCCT GATTGCCGCA TCCGCACTCT GCATTTC AAC
    51 CGCTTCGGCA CATCCTGCCA GCGAACCGTC CACCAAAAC GAAACCGCTA

```

10

```

1  MKHILPLIAA  SALCISTASA  HPASEPSTQN  ETAMTTHTLT  SKYSFDETVS
51 RLETAISKSG  MDIFAVIDHQ  EAARRNGLTM  QPAKVIVFGT  PKAGTPLMVK
101 DPAFALQLPL  RVLVTETDGG  VRAAYTDTRA  LIAGSRIGFD  EVANTLANAE
151 KLIQTVGE*

```

15

ORF97 shows 88.7% identity over a 159aa overlap with an ORF (ORF97a) from strain A of *N.*

20

25

30

40

45

50

55

```

                                10      20      30      40      50      60
orf97a.pep  MXHILPLXXASALCISTASXHPASEPQTQNETAMTTHTLTSKYSFDETVSRLETAIKSKG
             | | | | | | | | | | | | | | | | | | | | | | | | | | | |
orf97-1     MKHILPLIAASALCISTASAHPASEPSTQNETAMTTHTLTSKYSFDETVSRLETAIKSKG

```

-248-

		10	20	30	40	50	60
		70	80	90	100	110	120
5	orf97a.pep	MDIFAVIDHQEAAARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAFALQLPLRVXVTETDGK					
	orf97-1	MDIFAVIDHQEAAARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAFALQLPLRVLTETDGK					
		70	80	90	100	110	120
10	orf97a.pep	VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTIGEX					
	orf97-1	VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTVGEX					
		130	140	150	160		

15 Homology with a predicted ORF from *N.gonorrhoeae*

ORF97 shows 88.1% identity over a 159aa overlap with a predicted ORF (ORF97.ng) from *N. gonorrhoeae*:

	orf97.pep	MKHILPLIAASALCISTASAHASEPSTQNETAMITHTLISKYSFGXXXXXXAIKSKG	60
20	orf97ng	MKHILPPIAASAFICISTASAHAGKPPTQNETAMTTHTLTSKYSFDETVSRLETAIKSKG	60
	orf97.pep	MDIFAVIDHQEAAARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAFALQLPLRVLTETDGK	120
25	orf97ng	MDIFAVIDHQEAAARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAFALQLPLRVLTETDGK	120
	orf97.pep	VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTVGE	159
	orf97ng	VRTAYTDTRALIVGSRISFDEVANTLANAEKLIQKTVGE	159

The complete length ORF97ng nucleotide sequence <SEQ ID 361> is predicted to encode a protein having amino acid sequence <SEQ ID 362>:

```

1  MKHILPPIAA  SAFCISTASA  HPAGKPPTQN  ETAMTTHTLT  SKYSFDETVS
51  RLETAIKSKG  MDIFAVIDHQ  EAARRNGLTM  QPAKVIVFGT  PKAGTPLMVK
101 DPAFALQLPL  RVLVTETDGK  VRTAYTDTRA  LIVGSRISFD  EVANTLANAE
151 KLIQKTVGE*
```

## 35 Further work revealed the complete nucleotide sequence &lt;SEQ ID 363&gt;:

```

1  ATGAAACACA  TACTCCCcct  gatcgccgca  TccgcactCT  GCATTTCAC
51  CGCTTCGGCA  CACCCTGCCG  GCAAACCGCC  CACCCAAAAC  GAAACCGCTA
101 TGACCACGCA  CACCCTCACC  TCGAAATACA  GTTTGACGA  AACCGTCAGC
151 CGCCTTGAAA  CCGCCATAAA  AAGCAAAGGG  ATGGACATT  TTGCCGTCAT
40 201 CGACCATCAG  GAAGCGGCAC  GCCGAAACGG  CCTGACCATG  CAGCCGGCAA
251 AAGTCATCGT  CTTCCGGCAG  CCCAAGGCCG  GTACGCCgct  GATGGTCAAA
301 GACCCCGCCT  TCGCCCTGCA  ACTGCCCTG  CGCGTCCTCG  TTACCGAAAC
351 GGACGGCAAA  GTACGCACCG  CCTATACCGA  TACGCGCGCC  CTCATCGTCG
45 401 GCAGCCGCAT  CAGTTTCGAC  GAAGTGGCAA  ACACCTTGGC  AAACGCCGAA
451 AAAGTATAC  AAAAAACCGT  AGGCGAATAA
```

This corresponds to the amino acid sequence <SEQ ID 364; ORF97ng-1>:

```

1  MKHILPLIAA  SALCISTASA  HPAGKPPTQN  ETAMTTHTLT  SKYSFDETVS
51  RLETAIKSKG  MDIFAVIDHQ  EAARRNGLTM  QPAKVIVFGT  PKAGTPLMVK
50 101 DPAFALQLPL  RVLVTETDGK  VRTAYTDTRA  LIVGSRISFD  EVANTLANAE
151 KLIQKTVGE*
```

ORF97ng-1 and ORF97-1 show 96.2% identity in 159 aa overlap:

		10	20	30	40	50	60
	orf97-1.pep	MKHILPLIAASALCISTASAHASEPSTQNETAMTTHTLTSKYSFDETVSRLETAIKSKG					
55	orf97ng-1	MKHILPLIAASALCISTASAHAGKPPTQNETAMTTHTLTSKYSFDETVSRLETAIKSKG					
		10	20	30	40	50	60

		70	80	90	100	110	120
5	orf97-1.pep	MDIFAVIDHQAARRNGLTMQPAKVI	VFGTPKAGTPLMVKDPAFALQLPLRV	LVLTETDGK			
	orf97ng-1	MDIFAVIDHQAARRNGLTMQPAKVI	VFGTPKAGTPLMVKDPAFALQLPLRV	LVLTETDGK			
		70	80	90	100	110	120
10	orf97-1.pep	VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKT	VGEX				
	orf97ng-1	VRTAYTDTRALIVGSRISFDEVANTLANAEKLIQKT	VGEX				
		130	140	150	160		

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF97-1 (15.3kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figures 12A & 12B show, respectively, the results of affinity purification of the GST-fusion and His-fusion proteins. Purified GST-fusion protein was used to immunise mice, whose sera were used for Western Blot (Figure 12C), ELISA (positive result), and FACS analysis (Figure 12D). These experiments confirm that ORF97-1 is a surface-exposed protein, and that it is a useful immunogen.

Figure 12E shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF97-1.

#### Example 43

The following DNA, believed to be complete, sequence was identified in *N.meningitidis* <SEQ ID 365>:

```

1  ATGGCTTTTA TTACGCGCTT ATTCAAAAGC AGTAAATGGC TGATTGTGCC
51  GCTGATGCTC CCCGCTTTTC AGAATGTGGC GCGGAGGGG ATAGATGTGA
101 GCGTGCCGA AGCGAGGATA ACCGACGGCG GGCAGCTTTC CATCAGCAGC
151 CGCTTCCAAA CCGAGCTGCC CGACAGCTC CAACAGGCGT TCGCGCGGGg
201 CGTGCGGCTC AACTTTACCT TAAGCTGGCA GCTTCCGCC CCGATAATCG
251 CTTCTTATCG GTTTAAATTG GGGCAACTGA TTGGCGATGA CGACaATATT
301 GACTACAAAC TGAGTTTCCA TCCGCTGACc AaACGCTACC GCGTTACCGT
351 CGgCGCGTTT TCGACAGACT ACGACACCTT GGATGCGGCA TTGCGCGCGA
401 CCGGCGCGGT TGCCAACCTG AAAGTCCCTGA ACAAAGGCGC GCTGTCCGGT
35  451 GCGGAAGCAG GGGAAACCAA GGCGGAAATC CGCCTGACGC TGTCCACTTC
501 AAAACTGCCC AAGCCTTTTC AAATCAATGC ATTGACTTCT CAAAACCTGGC
551 ATTTGGATTG GGTGTTGAAA CCTCTAAACA TCATCGGGAA CAAATAA

```

This corresponds to the amino acid sequence <SEQ ID 366; ORF106>:

```

1  MAFITRLFKS SKWLIVPLML PAFQNVAAEG IDVSRAEARI TDGGQLSISS
51  RFQTELPDQL QQALRRGVPL NETLSWQLSA PIIASYRFLK GQLIGDDDDNI
101 DYKLSFHPLT KRYRVTVGAF STDYDTLDA LRTAGAVANW KVLNKGALSG
151 AEAGETKAEI RLTLSTSKLP KPFQINALTS QNWLHDSGWK PLNIIGNK*

```

Further work revealed the following DNA sequence <SEQ ID 367>:

```

1  ATGGCTTTTA TTACGCGCTT ATTCAAAAGC AGTAAATGGC TGATTGTGCC
51  GCTGATGCTC CCCGCTTTTC AGAATGTGGC GCGGAGGGG ATAGATGTGA

```

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55

```

101 GCCGTGCCGA AGCGAGGATA ACCGACGGCG GGCAGCTTTC CATCAGCAGC
151 CGCTTCCAAA CCGAGCTGCC CGACCAGCTC CAACAGGCGT TGCGCCGGGG
201 CGTGCCGCTC AACTTTACCT TAAGCTGGCA GCTTTCCGCC CCGATAATCG
251 CTTCTTATCG GTTTAAATTG GGGCAACTGA TTGGCGATGA CGACAATATT
301 GACTACAAAC TGAGTTTCCA TCCGCTGACC AACCCTACC GCGTTACCGT
351 CGGCGCGTTT TCGACAGACT ACGACACCTT GGATGCGGCA TTGCGCGCGA
401 CCGGCGCGGT TGCCAACTGG AAAGTCCTGA ACAAAGGCGC GCTGTCCGGT
451 GCGGAAGCAG GGGAAACCAA GGCGGAAATC CGCCTGACGC TGTCCACTTC
501 AAAACTGCCC AAGCCTTTC AAATCAATGC ATTGACTTCT CAAAACTGGC
551 ATTTGGATTG GGGTTGGAAA CCTCTAAACA TCATCGGGAA CAAATAA

```

This corresponds to the amino acid sequence <SEQ ID 368; ORF106-1>:

15  
20  
25  
30  
35  
40  
45  
50  
55

```

1 MAFITRLFKS SKWLIVPLML PAFQNVAAEG IDVSRAEARI TDGGQLSISS
51 RFQTELPDQL QQALRRGVPL NFTLSWQLSA PIASRYFKL GQLIGDDDDNI
101 DYKLSFHPLT NRYRVTVGAF STDYDTLDA LRTGAVANW KVLNKGALSG
151 AEAGETKAEI RLTLSTSKLP KPFQINALTS QNWHLD SGWK PLNIIGNK*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF106 shows 87.4% identity over a 199aa overlap with an ORF (ORF106a) from strain A of *N.*

*meningitidis*:

20  
25  
30  
35  
40

```

      10      20      30      40      50      59
orf106.pep MAFITRLFKSSK-WLIVPLMLPAFQNVAAEGIDVSRAEARITDGGQLSISSRFQTELPDQ
          ||||| ||| : || : || : ||||| ||||| ||||| |||||
orf106a    MAFITRLFKSIKQLVLLPMLSVLPDAAAEGIDVSRAEARIXDGGQLSXXSRFQTELPDQ
          10      20      30      40      50      60

      60      70      80      90      100     110     119
orf106.pep LQQALRRGVPLNFTLSWQLSAPIIASRYFKLGQLIGDDDDNIDYKLSFHPLTKRYRVTVGA
          || | ||| || || ||||| ||||| ||||| ||||| : |||||
orf106a    LQXAXXRGVXLNXTLXWQLSAPIIASRYFXLGLGDDDXIDYKLSFHPLTNRYRVTVGA
          70      80      90      100     110     120

      120     130     140     150     160     170     179
orf106.pep FSTDYDTLDAALRATGAVANWKVLNKGALSGAEAGETKAEIRLTLSTSKLPKPFQINALT
          || | ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf106a    FSTXYDTLDAALRATGAVANWKVLNKGALSGAEAGETKAEIRLTLSTSKLPKPFQINALT
          130     140     150     160     170     180

      180     190     199
orf106.pep SQNWHLD SGWKPLNIIGNKX
          ||||| ||||| |||||
orf106a    SQNWHLD SGWKPLNIIGNKX
          190     200

```

Due to the K→N substitution at residue 111, the homology between ORF106a and ORF106-1 is 87.9% over the same 199 aa overlap.

45 The complete length ORF106a nucleotide sequence <SEQ ID 369> is:

50  
55

```

1 ATGGCTTTTA TTACGCGCTT ATTCAAAAGC ATTAACAAT GGCTTGTGCT
51 GCTGCCGATG CTTTCCGTTT TGCCGGACGC GCGGCGGAG GGGATAGATG
101 TGAGCCGCGC CGAAGCGAGG ATAANCGACG GCGGGCAGCT TTCCATNAGN
151 AGCCGCTTCC AAACCGAGCT GCCCGACCAG CTCCAANNNG CGNNGNGCCG
201 GGGCGTGNCG CTCAACTNTA CCTTAAGNTG GCAGCTTTCC GCCCGATAA
251 TCGCTTCTTA TCGGTTTNA TTTGGGCAAC TGATTGGCGA TGACGACNAT
301 ATTGACTACA AACTGAGTTT CCATCCGCTG ACCAACCCT ACCGCGTTAC
351 CGTCGGCGCG TTTTCGACAG ANTACGACAC CTTGGATGCG GCATTGCGCG
401 CGACCGGCGC GGTGCCCAAC TGGAAAGTCC TGAACAAAG CGCGCTGTCC
451 GGTGCGGAAG CAGGGGAAAC CAAGGCGGAA ATCCGCCTGA CGCTGTCCAC
501 TTCAAAACTG CCCAAGCCTT TTCAAATCAA TGCATTGACT TCTCAAAACT

```



551 GGCATTGGA TTCGGTTGG AACCTCTAA ACATCATCGG GAACAAATAA

This encodes a protein having amino acid sequence <SEQ ID 370>:

1 MAFITRLFKS IKQWLVLPM LSVLPDAAAE GIDVSRAEAR IXDGGQLSXX  
51 SRFQTELPDQ LQXAXXRGVX LNXTLXWQLS APIIASYRFX LGQLIGDDDX  
101 IDYKLSFHPL TNRYRVTVGA FSTXYDTLDA ALRATGAVAN WKVLNKGALS  
151 GAEAGETKAE IRLTLSTSKL PKPFQINALT SQNWHLDSGW KPLNIIGNK\*

# Homology with a predicted ORF from *N.gonorrhoeae*

ORF106 shows 90.5% identity over a 199aa overlap with a predicted ORF (ORF106.ng) from *N.*

10 *gonorrhoeae*:

orf106.pep	MAFITRLFKSSK-WLIVPLMLPAFQNVAAEGIDVSRAEARITDGGQLSISSRFQTELPDQ	59
orf106ng	MAFITRLFKSIKQWLVLPLSVLPDAAAEAGIAATRAEARITDGGRLSISSRFQTELPDQ	60
orf106.pep	LQQALRRGVPLNFTLSWQLSAPIIASYRFXKLGLIGDDDNIDYKLSFHPLTKRYRVTVGA	119
orf106ng	LQQALRRGVPLNFTLSWQLSAPTIASYRFXKLGLIGDDDNIDYKLSFHPLTNRYRVTVGA	120
orf106.pep	FSTDYDTLDAALRATGAVANWKVLNKGALSGAEAGETKAEIRLTLSTSKLKPFPQINALT	179
orf106ng	FSTDYDTLDAALRATGAVANWKVLNKGALSGAEAGETKAEIRLTLSTSKLKPFPQINALT	180
orf106.pep	SQNWHLDSGWKPLNIIGNK	198
orf106ng	SQNWHLDSGWKPLNIIGNK	199

Due to the K→N substitution at residue 111, the homology between ORF106ng and ORF106-1 is 91.0% over the same 199 aa overlap.

The complete length ORF106ng nucleotide sequence <SEQ ID 371> is:

1 ATGGCTTTTA TTACGCGCTT ATTCAAAAGC ATTAAACAAT GGCTTGTGCT  
51 GTTGCCGATA CTCTCCGTTT TGCCGGACGC GCGGCGGAG GGCATTGCCG  
101 CGACCCGCGC CGAAGCGAGG ATAACCGACG GCGGGCGGCT TTCCATCAGC  
151 AGCGGCTTCC AAACCGAGCT GCCCGACCAG CTCCAACAGG CGTTGCGCCG  
201 GGGCGTACCG CTCAACTTTA CCTTAAGCTG GCAGCTTTCC GCCCGACAA  
251 TCGCTTCTTA TCGGTTTAAA TTGGGGCAAC TGATTGGCGA TGACGACAAT  
301 ATTGACTACA AACTAAGTTT CCATCCGCTG ACCAACCCTT ACCGCGTTAC  
351 CGTCGGCGCA TTTTCCACCG ATTACGACAC TTTGGATGCG GCATTGCGCG  
401 CGACCGGCGC GGTGCCCAAC TGGAAAGTCC TGAACAAAGG CGCGTTGTCC  
451 GGTGCGGAAG CAGGGGAAAC CAAGGCGGAA ATCCGCCTGA CGCTGTCCAC  
501 TTCAAAACTG CCAAGCCTT TCCAATCAA CGCATTGACT TCTCAAAACT  
551 GGCATTGGA TTCGGTTGG AACCTCTAA ACATCATCGG GAACAAATAA

This encodes a protein having amino acid sequence <SEQ ID 372>:

1 MAFITRLFKS IKQWLVLPI LSVLPDAAAE GIAATRAEAR ITDGGRLSIS  
51 SRFQTELPDQ LQQALRRGV LNFILSWQLS APTIASYRFX LGQLIGDDDN  
101 IDYKLSFHPL TNRYRVTVGA FSTDYDTLDA ALRATGAVAN WKVLNKGALS  
151 GAEAGETKAE IRLTLSTSKL PKPFQINALT SQNWHLDSGW KPLNIIGNK\*

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF106-1 (18kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 13A shows the results of affinity purification of the His-fusion protein, and Figure 13B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for FACS analysis (Figure 13C) These experiments confirm that ORF106-1 is a surface-exposed protein, and that it is a useful immunogen.

#### Example 44

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 373>:

```

10      1  ATGGACACAA AAGAAATCCT CGG.TACGCG GcAGGcTCGA TCGGCAGCGC
      51  GGTTTTAGCC GTCATCATCc TGCCGCTGCT GTCGTGGTAT TTCCCCGCCG
     101  ACGACATCGG GCGCATCGTG CTGATGCAGA CGGCGCGGG GCTgACGGTG
     151  TCGGTGTGTG GCCTCGGGCT GGATCAGGCA TACGTCCGCG AATACTATGC
     201  CACCGCCGAC AAAGACAcCT TGTTCAAAC CCTGTTCCTG CCGCCGCTGC
     251  TGTCTGCCGC CGGATAGCC GCCCTGCTGC TTTCCCGCCC GTCCCTGCCG
     301  TCTGAAATCC TGTTTTCACT CGACGATGCC gCCGCCGGCa TCGGGCTGGT
     351  GCTGTTTGAA CtGAGCTTCC TGCCCATCCG cTTTCTCTTA CTGGTTTTCG
     401  GTATGGAAGG ACGCGCCcTT GCCTTTTCGT CCGCGCAACT CGTGCCcAAG
     451  CTCGCCATCC TGCTGCTG.T GCCGCTGACG GTCGGGCTGC TGCACTTTCC
     501  AGCGAACACC GCCGTCCTGA CCGCGGTTTA CGCGCTGGCA AACCTTGCCG
     551  CCGCCGCCCTT TTTGCTGTTT CAAACCGAT GCCGTCTGAA GGCCGTCCGG
     601  CACGCACCGT TTTGCCCGCG CGTCTGCAC CGGGGG.TGC GCTACGGCAT
     651  ACCGATCGCA CTGAGCAGCA TCGCCTATTG GGGGCTGGCA TCCCGCGACC
     701  GTTTGTTTCT GAAAAAATAT GCCGGCCTGG AACAGCTCGG CGTTTATTTCG
     751  ATGGGTATTT CGTTCGGCGG GCGGCATTA TTGTCCAAA GCATCTTTTC
     801  AACGGTCTGG ACACCGTATA TTTCCGCGC AATCGAAGAA AACGCCCCGC
     851  CCGCTCGCCT CTCGGCAACG GCAGAATCCG CCGCCGCCCT GCTTGCCCTC
     901  GCCCTCTGC. TGACCGGCAT TTTCTGCCCC CTTGCCCTCC TCCTGCTGCC
     951  GGAAAACTAC GCCGCCGTCC GGTATATCGT CGTATCGTGT ATG.TGCCCG
    1001  CGCTGTTTTG CACGCTGGCG GAAATCAGCG GCATCGGTTT GAACGTCGTT
    1051  CGCAAAACGC GCCGATCGC GCTCGCCACC TTGGGCGCGC TGGCGGCAAA
    1101  CCTGCTGCTG CTGGGGCTTG ACCGTGCCGT ACCGCGAGG CCGCC.GGCG
    1151  CGGCGGTGCG CTGTGCCGCC TCATTCTGGC TGTTTTTTGC CTTCAAGACC
    1201  GAAAGCTCyT GCGCCTGTG GCAGCCGCTC AAACGCTGC CGCTTTATCT
    1251  GCACACATTG TTCTGCCTGA CCTCCTCGGC GGCCTACACC TGCTTCGGCA
    1301  CGCCGGCAAA CTATCCCTG TTTGCCGGCG TATGGCGGCG ATATCTGGCA
    1351  GGCTGCATCC TCGCCACCG GAAAGATTG CACAACTGT TTCATTATTT
    1401  GAAAAACAA GGTTTCCAT TATGA

```

This corresponds to the amino acid sequence <SEQ ID 374; ORF10>:

```

40      1  MDTKEILXYA AGSIGSAVLA VIILPLLSWY FPADDIGRIV LMQTAAGLTV
      51  SVLCLGLDQA YVREYYATAD KDTLFKTLFL PLLSAAAIA ALLLSRPSLP
     101  SEILFSLDDA AAGIGLVLE LSLPIRFL LVLMEGRAL AFSSAQLVPK
     151  LAILLXPLT VGLLHFPANT AVLTAVYALA NLAAAFLLF QNRCRLKAVR
     201  HAPFSPAVLH RGXYGIPIA LSSIAYWGLA SADRLFLKKY AGLEQLGVYS
     251  MGISFGGAAL LFQSFSTVW TPYIFRAIEE NAPPARLSAT AESAAALLAS
     301  ALCXTGIFSP LASLLLPENY AAVRFIVVSC MXPPLEFCTLA EISGIGLNVV
     351  RKTRPIALAT LGALAAANLL LGLDRAVPAR PXGAAVACAA SFWLFFAFKT
     401  ESSCRLWQPL KRLPLYLHTL FCLTSSAAYT CFGTPANYPL FAGVWAAYLA
     451  GCILRHRKDL HKLFHYLKKQ GFPL*

```

50 Further sequence analysis revealed the complete DNA sequence<SEQ ID 375> to be:

```

      1  ATGGACACAA AAGAAATCCT CGGCTACGCG GCAGGCTCGA TCGGCAGCGC
      51  GGTTTTAGCC GTCATCATCC TGCCGCTGCT GTCGTGGTAT TTCCCCGCCG
     101  ACGACATCGG GCGCATCGTG CTGATGCAGA CGGCGCGGG GCTGACGGTG

```

5  
10  
15  
20  
25

```

151 TCGGTGTTGT GCCTCGGGCT GGATCAGGCA TACGTCCGCG AATACTATGC
201 CACCGCCGAC AAAGACACCT TGTTCAAAAC CCTGTTCTCTG CCGCCGCTGC
251 TGTCTGCCGC CGCGATAGCC GCCCTGCTGC TTTCCCGCCC GTCCCTGCCG
301 TCTGAAATCC TGTTTTCACT CGACGATGCC GCCGCCGGCA TCGGGCTGGT
351 GCTGTTTGAA CTGAGCTTCC TGCCCATCCG CTTTCTCTTA CTGGTTTGC
401 GTATGGAAGG ACGCGCCCTT GCCTTTTCGT CCGCGCAACT CGTGCCCAAG
451 CTCGCCATCC TGCTGTGCT GCCGCTGACG GTCGGGCTGC TGCACTTTCC
501 AGCGAACACC GCCGTCTGA CCGCCGTTA CGCGCTGGCA AACCTTGCCG
551 CCGCCGCCCTT TTGCTGTTT CAAAACCGAT GCCGTCTGAA GGCCGTCCGG
601 CACGCACCGT TTTCGCCCGC CGTCTGCAC CGGGGGCTGC GCTACGGCAT
651 ACCGATCGCA CTGAGCAGCA TCGCCTATTG GGGGCTGGCA TCCGCCGACC
701 GTTTGTTCCT GAAAAAATAT GCCGGCCTGG AACAGCTCGG CGTTTATTCTG
751 ATGGGTATTT CGTTCGGCGG GGCGGCATTA TTGTTCCAA GCATCTTTTC
801 AACGGTCTGG ACACCGTATA TTTTCCGCGC AATCGAAGAA AACGCCCGGC
851 CCGCCCGCCTT CTCGGCAACG GCAGAATCCG CCGCCGCCCT GCTTGCTCC
901 GCCCTCTGCC TGACCGGCAT TTTCTCGCCC CTGCGCTCCC TCCTGCTGCC
951 GGAAGACTAC GCCGCCGTCG GGTATATCGT CGTATCGTGT ATGCTGCCGC
1001 CGCTGTTTTC CACGCTGGCG GAAATCAGCG GCATCGGTTT GAACGTCGTC
1051 CGCAAAACGC GCCGATCGC GCTCGCCACC TTGGGCGCGC TGGCGGCAAA
1101 CCTGCTGCTG CTGGGGCTTG CCGTGCCGTC CGGCGGCGCG CGCGGCGCGG
1151 CGGTGCTGCTG TGCCGCCTCA TTCTGGCTGT TTTTGCCTT CAAGACCGAA
1201 AGCTCCTGCC GCCTGTGGCA GCCGCTCAA CGCCTGCCG TTTATCTGCA
1251 CACATTGTTT TGCTGACCT CCTCGGCGGC CTACACCTGC TTCGGCAGCG
1301 CGGCAACTA TCCCTGTTT GCCGGCGTAT GGGCGGCATA TCTGGCAGGC
1351 TGCATCTGC GCCACCGGAA AGATTGTCAC AAAGTGTTC ATTATTGAA
1401 AAAACAAGGT TTCCATTAT GA

```

This corresponds to the amino acid sequence <SEQ ID 376; ORF10-1>:

30  
35

```

1 MDTKELGYA AGSIGSAVLA VIILPLLSWY FPADDIGRIV LMQTAAGLTV
51 SVLCLGLDQA YVREYYATAD KDTLFKTLFL PPLLSAAAIA ALLLSRPSLP
101 SEILFSLDDA AAGIGLVLE LSELPFRLL LVLMEGRAL AFSSAQLVPK
151 LAIIIIIIPLT VGLLHFPANT AVLTAVYALA NLAAAFLLF QNRCRLKAVR
201 HAPFSPAVLH RGLRYGIPIA LSSIAWGLA SADRLFLKKY AGLEQLGVYS
251 MGISFGGAAL LFQSFSTVW TPYIFRAIEE NAPPARLSAT AESAALLAS
301 ALCLTGIFSP LASLLLPENY AAVRFIVVSC MLPPLFCTLA EISGIGLNVV
351 RKTRPIALAT LGALANLLL LGLAVPSGGA RGAAVACAAS FWLFFAEKTE
401 SSCRLWQPLK RLPLYLHTLF CLTSSAAYTC FGTPANYPLF AGVWAAYLAG
451 CILRHRKDLH KLFHYLKKQG FPL*

```

Computer analysis of this amino acid sequence gave the following results:

#### Prediction

40 ORF10-1 is predicted to be the precursor of an integral membrane protein, since it comprises several (12-13) potential transmembrane segments, and a probable cleavable signal peptide

#### Homology with EpsM from *Streptococcus thermophilus* (accession number U40830).

ORF10 shows homology with the epsM gene of *S. thermophilus*, which encodes a protein of a size similar to ORF10 and is involved in expolysaccharide synthesis. Other homologies are with  
45 prokaryotic membrane proteins:

Identities = (25%)

50 Query: 213 LRYGIPLALSSLAYWGLASADRLFLKKYAGLEQLGVYSMGISFGGAALLQSIFSTVW 270  
L Y +PL SS+ +W L ++ R F+ + G G+ ++ + +IF+ W  
Sbjct: 210 LYYALPLIPSSILWLLNASSRYFVLFFLGAGANGLAVATKIPSIISIFTITQAW 267

Identities = 15/57 (26%), Positives = 31/57 (54%)

55 Query: 7 LGYAAGSIGSAVLAVIILPLLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQAYVR 63  
L + G++GS +L +++PL ++ + G L QT A L + ++ + + A +R  
Sbjct: 12 LVFTIGNLGSKLLVFLVPLYTYAMTPQEYGMADLYQTANLLLPLITMNVFDTLR 68

Identities = 16/96 (16%), Positives = 36/96 (37%)

Query: 307 IFSPLASLLPENYAARFTVVSCLPPLFYTLTEISGIGLNVVRKTRPIXXXXXXXXX 366  
 + P+ ++ +YA+ V ML LF + ++ G ++T+ +  
 Sbjct: 305 VLKPIVEKVSSDYASSWQYVPFMLSMLFSSFSDFGTNYIAAKQTKGVFMTSIYGTIV 364

Homology with a predicted ORF from *N.meningitidis* (strain A)ORF10 shows 95.4% identity over a 475aa overlap with an ORF (ORF10a) from strain A of *N.*10 *meningitidis*:

		10	20	30	40	50	60
orf10.pep		MDTKEILXYAAGSIGSAVLAVIILPPLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA					
orf10a		MDTKEILGYAAGSIGSAVLAVIILPPLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA					
		10	20	30	40	50	60
orf10.pep		YVREYYATADKDTLFTKTLFLPPLSAAAIALLSRPSLPSEILFSLDDAAAGIGLVLFE					
orf10a		YVREYYAADKDTLFTKTLFLPPLSAAAIALLSRPSLPSEILFSLDDAAAGIGLVLFE					
		70	80	90	100	110	120
orf10.pep		LSFLPIRFLLLVLRMEGRALAFSSAQLVSKLAILLLPLTVGLLHFPANTAVLTAVYALA					
orf10a		LSFLPIRFLLLVLRMEGRALAFSSAQLVSKLAILLLPLTVGLLHFPANTAVLTAVYALA					
		130	140	150	160	170	180
orf10.pep		NLAAAFLLFQNRCLKAVRHAPFSPAVLHRGXRYGIPIALSSIAIWGLASADRLFLKKY					
orf10a		NLAAAFLLFQNRCLKAVRRAPFSSAVLHRGLRYGIPIALSSIAIWGLASADRLFLKKY					
		190	200	210	220	230	240
orf10.pep		AGLEQLGVSMGISFSGAALLFQSIFSTVWTPYIFRAIEENAPPARLSATAESAAALLAS					
orf10a		AGLEQLGVSMGISFSGAALLFQSIFSTVWTPYIFRAIEANAPPARLSATAESAAALLAS					
		250	260	270	280	290	300
orf10.pep		ALCXTGIFSPLASLLLPENYAARFIVVSCMXPLFCTLAEISGIGLNVVRKTRPIALAT					
orf10a		ALCLTGIFSPLASLLLPENYAARFIVVSCMLPPLFCTLVEISGIGLNVVRKTRPIALAT					
		310	320	330	340	350	360
orf10.pep		LGALAANLLLLGLDRAVPAR-PXGAAVACAASFVLFVFKTESSCRLWQPLKRLPLYLHT					
orf10a		LGALAANLLLLGL--AVPSGGARGAACAASFVLFVFKTESSCRLWQPLKRLPLYMHT					
		370	380	390	400	410	419
orf10.pep		LFCLTSSAAYTCFGTPANYPLFAGVWAAYLAGCILRHRKDLHKLPHYLLKKQGFPLX					
orf10a		LFCLASSAAYTCFGTPANYPLFAGVWAVYLAGCILRHRKDLHKLPHYLLKKQGFPLX					
		420	430	440	450	460	470

The complete length ORF10a nucleotide sequence &lt;SEQ ID 377&gt; is:

60 1 ATGGACACAA AAGAAATCCT CGGCTACGCG GCAGGCTCGA TCGGCAGCGC  
 51 GGTTTtagcc GTCATCATCC TGCCGCTGCT GTCGTGGTAT TTCCCTGCCG  
 101 ACGACATCGG ACGCATCGTG CTGATGCAGA CGGCGGCGGG GCTGACGGTG  
 151 TCGGTGTTGT GCCTCGGGCT GGATCAGGCA TACGTCCGCG AATACTATGC  
 201 CGCCGCCGAC AAAGACACTT TGTCAAAC CCTGTTCTG CCGCCGCTGC  
 251 TGTCTGCCG CGCGATAGCC GCCCTGCTG TTTCCGCC ATCCCTGCCG

5  
10  
15  
20

301 TCTGAAATCC TGTTCGCT CGACGATGCC GCCGCCGGCA TCGGGCTGGT  
351 GCTGTTTGA CTGAGCTTCC TGCCCATCCG CTTTCTCTTA CTGGTTTTCG  
401 GTATGGAAGG ACGCGCCCTT GCCTTTTCGT CCGCGCAACT CGTGTCCAAG  
451 CTCGCCATCC TGCTGCTGCT GCCGCTGACG GTCGGGCTGC TGCACCTTCC  
501 GCGGAACACC GCCGTCCTGA CCGCGCTTTA CGCGCTGGCA AACCTTGCCG  
551 CCGCCGCCCTT TTTGCTGTTT CAAAACCGAT GCCGCTGAA GGCCGTCGCG  
601 CGCGCACCGT TTTTCATCCG CGTCCTGCAT CGCGGCCTGC GCTACGGCAT  
651 ACCGATCGCA CTAAGCAGCA TCGCCTATTG GGGGCTGGCA TCCGCCGACC  
701 GTTTGTTTCT GAAAAAATAT GCCGGCCTAG AACAGCTCGG CGTTTATTCG  
751 ATGGGTATTT CGTTCGGCGG AGCGGCATTA TTGTTCCAAA GCATCTTTTC  
801 AACGGTCTGG ACACCGTATA TTTTCCGCGC AATCGAAGCA AACGCCCGCG  
851 CCGCCCGCCT CTCGGCAACG GCAGAATCCG CCGCCGCCCT GCTTGCCTCC  
901 GCCCTCTGCC TGACCGGCAT TTTCTCGCCC CTGCGCTCCC TCCTGCTGCC  
951 GGAAAACTAC GCCGCCGCTC GGTATTATCGT CGTATCGTGT ATGCTGCCTC  
1001 CGCTGTTTTG CACGCTGGTA GAAATCAGCG GCATCGGTTT GAACGTCGTC  
1051 CGAAAAACAC GCCCGATCGC GCTCGCCACC TTGGGCGCGC TGGCGGCAAA  
1101 CCTGCTGCTG CTGGGGCTTG CCGTACCGTC CGCGCGCGCG CGCGGCGCGG  
1151 CGGTTGCCTG TGCCGCCTCA TTTTGGCTGT TTTTGTGTTT CAAGACCGAA  
1201 AGCTCCTGCC GCCTGTGGCA GCCGCTCAA CGCCTGCCGC TTATATATGA  
1251 CACATTGTTC TGCTGGCCT CCTCGGCGGC CTACACCTGC TTCGGCACTC  
1301 CGGCAACTA CCCCCTGTT GCCGGCGTAT GGGCGGTATA TCTGGCAGGC  
1351 TGCATCTGCG GCCACCGGAA AGATTGTCAC AAAGTGTTC ATTATTTGAA  
1401 AAAACAAGGT TTCCATTAT GA

This encodes a protein having amino acid sequence <SEQ ID 378>:

25  
30  
35

1 MDTKEILGYA AGSIGSAVLA VIILPLLSWY FPADDIGRIV LMQTAAGLTV  
51 SVLCLGLDQA YVREYYAAAD KDTLFKTLFL PPLLSAAAIA ALLSRPSLP  
101 SEILFSLDDA AAGIGLVLE LSFLPIRLL LVRMEGRAL AFSSAQLVSK  
151 LAILLLLPLT VGLLHFPANT AVLTAVYALA NLAAAFLF QNRCRLKAVR  
201 RAPFSSAVLH RGLRYGPIA LSSIAWGLA SADRLFLKKY AGLEQLGVYS  
251 MGISFGGAAL LQSFSTVW TPYIFRAIEA NAPPARLSAT AESAAALLAS  
301 ALCLTGIFSP LASLLLPENY AAVRFIVVSC MLPLFCTLV EISIGLNVV  
351 RKTRPIALAT LGALANLLL LGLAVPSGGA RGAAVACAAS FWLFFVFKTE  
401 SSCRLWQPLK RLPLYMHTLF CLASSAAYTC FGTPANYPLF AGVWAVYLAG  
451 CILRHRKDLH KLFHYLKKQG FPL\*

35 ORF10a and ORF10-1 show 95.4% identity in 475 aa overlap:

40  
45  
50  
55  
60  
65

	10	20	30	40	50	60
orf10-1.pep	MDTKEILXYAAGSIGSAVLAVIILPLLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA					
orf10a	MDTKEILGYAAGSIGSAVLAVIILPLLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA					
	70	80	90	100	110	120
orf10-1.pep	YVREYYATADKDTLFKTLFLPPLLSAAAIAALLSRPSLPSEILFSLDDAAAGIGLVLE					
orf10a	YVREYYAAADKDTLFKTLFLPPLLSAAAIAALLSRPSLPSEILFSLDDAAAGIGLVLE					
	130	140	150	160	170	180
orf10-1.pep	LSFLPIRLLLVLRMEGRALAFSSAQLVSKLAILLLLPLTVGLLHFPANTAVLTAVYALA					
orf10a	LSFLPIRLLLVLRMEGRALAFSSAQLVSKLAILLLLPLTVGLLHFPANTAVLTAVYALA					
	190	200	210	220	230	240
orf10-1.pep	NLAAAFLFQNRCRLKAVRHAPFSPAVLHRGXRYGPIALSSIAWGLASADRLFLKKY					
orf10a	NLAAAFLFQNRCRLKAVRRAPFSSAVLHRLRYGPIALSSIAWGLASADRLFLKKY					
	250	260	270	280	290	300
orf10-1.pep	AGLEQLGVYSMGISFGGAALLFQSIFSTVWTPYIFRAIEANAPPARLSATAESAAALLAS					
orf10a	AGLEQLGVYSMGISFGGAALLFQSIFSTVWTPYIFRAIEANAPPARLSATAESAAALLAS					
	310	320	330	340	350	360

5 orf10-1.pep ALCXTGIFSPLASLLLPENYA AVR FIVVSCMX PPLFCTLAEISGIGLNVVRKTRPIALAT  
 orf10a ALCXTGIFSPLASLLLPENYA AVR FIVVSCMLPPLFCTLVEISGIGLNVVRKTRPIALAT  
 310 320 330 340 350 360

10 orf10-1.pep LGALAAANLLLLGLDRAVPAR-PXGA AVACAASFWLFFAFKTESSCRLWQPLKRLPLYLHT  
 orf10a LGALAAANLLLLGL--AVPSGGARGA AVACAASFWLFFVFKTESSCRLWQPLKRLPLYMHT  
 370 380 390 400 410

15 orf10-1.pep 420 430 440 450 460 470  
 LFCLTSSAAYTCFGTPANYPLFAGVWAAYLAGCILRHRKDLHKL FHYLKKQGFPPLX  
 orf10a LFCLASSAAYTCFGTPANYPLFAGVWAAYLAGCILRHRKDLHKL FHYLKKQGFPPLX  
 420 430 440 450 460 470

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF10 shows 94.1% identity over a 475aa overlap with a predicted ORF (ORF10.ng) from *N.*

20 *gonorrhoeae*:

orf10ng.pep MDTKEILGYAAGSIGSAVLAVIILPLLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA 60  
 orf10nm MDTKEILXYAAGSIGSAVLAVIILPLLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA 60

25 orf10ng.pep YVREYYAAADKDTLFKTLFLPPLLSAAIAALLSRPSLPSEILFSLDDAAAGIGLVLFE 120  
 orf10nm YVREYYATADKDTLFKTLFLPPLLSAAIAALLSRPSLPSEILFSLDDAAAGIGLVLFE 120

30 orf10ng.pep LSFLPIRFLLLVLRMEGRALAFSSAQLVPKLAILLPLTVGLLHFPANTSVLTAVYALA 180  
 orf10nm LSFLPIRFLLLVLRMEGRALAFSSAQLVPKLAILLPLTVGLLHFPANTAVLTAVYALA 180

35 orf10ng.pep NLAAAFLFLQNRCLKAVRRAPFSPAVLHRGLRYGIPLALSSLAYWGLASADRLFLKKY 240  
 orf10nm NLAAAFLFLQNRCLKAVRHAPFSPAVLHRGXRYGIPLALSSIAYWGLASADRLFLKKY 240

40 orf10ng.pep AGLEQLGVYSMGISFGGAALLQSIFSTVWTPYIFRAIEENATPARLSATAESAAALLAS 300  
 orf10nm AGLEQLGVYSMGISFGGAALLQSIFSTVWTPYIFRAIEENAPPARLSATAESAAALLAS 300

45 orf10ng.pep ALCXTGIFSPLASLLLPENYA AVR FTVVSCMLPPLFYTLTEISGIGLNVVRKTRPIALAT 360  
 orf10nm ALCXTGIFSPLASLLLPENYA AVR FIVVSCMX PPLFCTLAEISGIGLNVVRKTRPIALAT 360

50 orf10ng.pep 370 380 390 400 410  
 LGALAAANLLLLGL--AVPSGGTGA AVACAASFWLFFVFKTESSCRLWQPLKRLPLYMHT  
 orf10nm LGALAAANLLLLGLDRAVPAR-PXGA AVACAASFWLFFAFKTESSCRLWQPLKRLPLYLHT  
 370 380 390 400 410

55 orf10ng.pep 420 430 440 450 460 470  
 LFCLASSAAYTCFGTPANYPLFAGVWAAYLAGCILRHRKNLHKL FHYLKKQGFPPLX  
 orf10nm LFCLTSSAAYTCFGTPANYPLFAGVWAAYLAGCILRHRKDLHKL FHYLKKQGFPPLX  
 420 430 440 450 460 470

The complete length ORF10ng nucleotide sequence <SEQ ID 379> is:

1 ATGGACACAA AAGAAATCCT CGGCTACGCG GCAGGCTCGA TCGGCAGCGC  
 51 GGTTTTAGCC GTCATCATCC TGCCGCTGCT GTCGTGGTAT TTCccgCCG  
 101 ACGACATCGG GCGCATCGTG CTGATGCAGA CGGCGGCGGG ACTGACGGTG  
 151 TCGGTATGT GCCTCGGGCT GGATCAGGCA TACGTCCGCG AATACTATGC  
 201 CGCCGCCGAC AAAGACACTT TGTTCAAAC CCTGTTCTTG CCGCCGCTGC  
 251 TGTTTTCCGC CGCGATAGCC GCCCTGCTGC TTTCCCGCCC GTCCCTGCCG  
 301 TCTGAAATCC TGTTTTCGCT CGACGATGCC GCCGCCGGCA TCGGGCTGGT  
 351 GCTGTTTGAA CTGAGCTTCC TGCCCATCCG CTTTCTCTTA CTGGTTTTCG

5  
10  
15  
20

```
401 GTATGGAAGG GCGCGCCCTT GCCTTTTCGT CCGCGCAACT CGTGCCCAAA
451 CTCGCCATTC TGCTGCTGTT GCCGCTGACG GTCGGGCTGC TGCACTTTCC
501 GCGGAACACC TCCGTCCTGA CCGCCGTTTA CCGGCTGGCA AACCTTGCCG
551 CCGCCGCTT TTTGCTGTTT CAAAACCGAT GCCGTCTGAA GGCCGTCCGG
601 CCGCGCCCGT TTTCGCCCGC CGTCCTGCAC CGGGGGCTGC GCTACGGCAT
651 ACCGCTCGCA CTGAGCAGCC TTGCCTATTG GGGGCTGGCA TCCGCCGACC
701 GTTTGTTTCT GAAAAAATAT GCGGGCCTGG AACAGCTCGG CGTTTATTTCG
751 ATGGGTATTT CGTTCGGCGG GGCGGCATTA TTGCTCCAAA GCATCTTTTC
801 AACGGTCTGG ACACCGTATA TTTTCCGTGC AATCGAAGAA AACGCCACGC
851 CCGCCGCTCT CCGGCAACG GCAGAATCCG CCGCCGCCCT GCTTGCCTCC
901 GCCTCTGCTC TGACCGGAAT TTTCTCGCCC CTCGCCTCCC TCCTGCTGCC
951 GGAAACTAC GCCGCCGTCC GGTTTACCGT CGTATCGTGT ATGCTGccgc
1001 cgctGTTTTA CACGCTGACC GAAATCAGCG GCATCGGTTT GAACGTCGTC
1051 CGCAAAACGC GTCCGATCGC GCTTGCCACC TTGGGCGCGC TGGCGGCAAA
1101 CCTGCTGCTG CTGGGGCTTG CCGTACCGTC CGGCGGCACG CGCGGCGCGG
1151 CGGTGCGCTG TGCCGCTCA TTCTGTTGT TTTTGTGTTT CAAGACAGAA
1201 AGCTCCTGCC GCCTGTGGCA GCCGCTCAA CGCTGCCGC TTTATATGCA
1251 CACATTGTTC TGCTGCGCT CCTCGCGGC CTACACCTGC TTCGGCACAC
1301 CCGCAACTA CCCcctgtt gccggcgtAT GGGCGGCATA TCTGGCAGGC
1351 TGCATCTGCG GCCACCGGAA AAATTGTCAC AACTGTGTTT ATTATTGAA
1401 AAAACAAGGT TTCCATTAT GA
```

This encodes a protein having amino acid sequence <SEQ ID 380>:

25  
30

```
1 MDTKEILGYA AGSIGSAVLA VIILPLLSWY FPADDIGRIV LMQTAAGLTV
51 SVLCIGLDQA YVREYYAAAD KDTLFKTLFL PPLLFSAIA ALLSRPSLP
101 SEILFSLDDA AAGIGLVLE LSELPPIRFL LVLRMEGRAL AFSSAQLVPK
151 LAIIIIIIPLT VGLLHFPANT SVLTAVYALA NLAAAFLLF QNRCRLKAVR
201 RAFFSPAVLH RGLRYGIPLA LSSLAYWGLA SADRLFLKKY AGLEQLGVYS
251 MGISFGGAAL LQSIFSTVW TPYIFRAIEE NATPARLSAT AESAAALLAS
301 ALCLTGIFSP LASLLLPENY AAVRFTVSC MLPPLFYTLT EISGIGLNVV
351 RKTRPIALAT LGALANLLL LGLAVPSGGT RGAAVACAAS FWLFFVFKTE
401 SSCRLWQPLK RPLYMHTLF CLASSAAYTC FGTPANYPLF AGVWAAAYLAG
451 CILRHRKNLH KLFHYLKKQG FPL*
```

ORF10ng and ORF10-1 show 96.4% identity in 473 aa overlap:

35  
40  
45  
50  
55  
60  
65

```
10 20 30 40 50 60
orf10-1.pep MDTKEILGYAAGSIGSAVLA VIILPLLSWYFPADDIGRIVLMQTAAGLTVSVLCIGLDQA
|||||
orf10ng-1 MDTKEILGYAAGSIGSAVLA VIILPLLSWYFPADDIGRIVLMQTAAGLTVSVLCIGLDQA
10 20 30 40 50 60

70 80 90 100 110 120
orf10-1.pep YVREYYATADKDTLFKTLFLPPLLSAAIAALLSRPSLPSEILFSLDDAAGIGLVLE
|||||:|||||:|||||:|||||:|||||:|||||
orf10ng-1 YVREYYAAADKDTLFKTLFLPPLLFSAIAALLSRPSLPSEILFSLDDAAGIGLVLE
70 80 90 100 110 120

130 140 150 160 170 180
orf10-1.pep LSFLPIRFLLLVLRMEGRALAFSSAQLVPKLAIILLLPLTVGLLHFPANTAVLTAVYALA
|||||
orf10ng-1 LSFLPIRFLLLVLRMEGRALAFSSAQLVPKLAIILLLPLTVGLLHFPANTSVLTAVALA
130 140 150 160 170 180

190 200 210 220 230 240
orf10-1.pep NLAAAFLLFQNRCRLKAVRHAPFSPAVLHRLRYGIPIALSSIAWGLASADRLFLKKY
|||||:|||||:|||||:|||||:|||||:|||||
orf10ng-1 NLAAAFLLFQNRCRLKAVRRAPFSPAVLHRLRYGIPLALSSLAYWGLASADRLFLKKY
190 200 210 220 230 240

250 260 270 280 290 300
orf10-1.pep AGLEQLGVYSMGISFGGAALLQSIFSTVWTPYIFRAIEENAPPARLSATAESAAALLAS
|||||:|||||:|||||:|||||:|||||:|||||
orf10ng-1 AGLEQLGVYSMGISFGGAALLQSIFSTVWTPYIFRAIEENATPARLSATAESAAALLAS
250 260 270 280 290 300

310 320 330 340 350 360
orf10-1.pep ALCLTGIFSPASLLLPENYAAVRFIVVSCMLPPLFCTLAIEISGIGLNVVRKTRPIALAT
|||||:|||||:|||||:|||||:|||||:|||||
```

	orf10ng-1	ALCLTGIFSP	LASLLPEN	YAAVRFTV	VSCMLPPL	FYTLTEIS	GIGLNVVR	KTRPIALAT
		310	320	330	340	350	360	
5	orf10-1.pep	LGALAANLLL	LGLAVPSG	GARGA	AAVACAA	FWLFFAFK	TESSCRLW	QPLKRLPLYLHTLF
	orf10ng-1	LGALAANLLL	LGLAVPSG	GTRGA	AAVACAA	FWLFFVFK	TESSCRLW	QPLKRLPLYMHTLF
		370	380	390	400	410	420	
10	orf10-1.pep	CLTSSAAYTC	FGTPANYPL	FAGVWAAY	LAGCILRHR	KDLHKLFH	YLLKKQGF	PLX
	orf10ng-1	CLASSAAYTC	FGTPANYPL	FAGVWAAY	LAGCILRHR	KNLHKLFH	YLLKKQGF	PLX
15		430	440	450	460	470		

Based on this analysis, including the presence of a putative leader peptide and several transmembrane segments and the presence of a leucine-zipper motif (4 Leu residues spaced by 6 aa, shown in bold), it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 45

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 381>:

1	..ATCCTGAAAC	CGCATAACCA	GCTTAAGGAA	GACATCCAAC	CTGATCCGGC
51	CGATCAAAAC	GCCTTGTC	CGAACC	GATGCG	CACAGAGC
101	CGGATGCGGA	AAATGCTGCC	GACAAGCAGC	CCGTTGCCGA	TAAAGCCGAC
151	GAGGTGTAAG	AAAAGGCGGG	CGAGCCGGA	CGGGAAGAGC	CGGACGGACA
201	GGCAGTGCGT	AAGAAAGCGC	TGACGGAAGA	GCGTGAACAA	ACCGTCAGGG
251	AAAAGCGCA	GAAGAAAGAT	GCCGAAACGG	TTAAATACA	AGCGGTAAAA
301	CCGTCTAAG	AAACAGAGAA	AAAAGCTTCA	AAAGAAGAGA	AAAAGGCGGC
351	GAAGGAAAA	GTTGCACCCA	AACCAACCCC	GGAACAAATC	CTCAACAGCG
401	GCAgCATCGA	AAAGCGCGC	AgTGCCGCGC	CCAAAGAAGT	GCAGAAAATG
451	AA.AACGTCC	GACAAGGCGG	AAGC.AACGC	ATTATCTGCA	AATGGGCGCG
501	TATGCCGACC	GTGAGAGCGC	GGAAGGCGAG	CGTGCCAAAC	TGGCAATCTT
551	GGGCATATCT	TCCAAGGTGG	TCGTTATCA	GGCGGGACAT	AAAACGCTTT
601	ACCGGGTGCA	AAGCGGCAAT	ATGTCTGCCG	ATGCGGTGA	

This corresponds to the amino acid sequence <SEQ ID 382; ORF65>:

1	..ILKPHNQLKE	DIQPDPAQN	ALSEPDAATE	AEQSDAENAA	DKQPVADKAD
51	EVEEKAGEPE	REEPDGQAVR	KKALTEEREQ	TVREKAQKKD	AETVKIQAVK
101	PSKETEEKAS	KEEKKAKEK	VAPKPTPEQI	LNSGSIEXAR	SAAAKEVQKM
151	XNVRQGSXR	IICKWARMPT	VRARKGSVPN	WQSWAYLPRW	SVIRRDIKRF
201	TGCKAAICLP	MR*			

Further work revealed the complete nucleotide sequence <SEQ ID 383>:

1	ATGTTTATGA	ACAAATTTTC	CCAATCCGGA	AAAGGTCTGT	CCGGTTTTTT
51	CTTCGGTTTG	ATACTGGCGA	CGGTCATTAT	TGCCGGTATT	TTGTTTTATC
101	TGAACACAGAG	CGGTCAAAT	GCGTTCAAAA	TCCCGGCTTC	GTCCAAGCAG
151	CCTGCAGAAA	CGGAAATCCT	GAAACCGAAA	AACCAGCCTA	AGGAAGACAT
201	CCAACCTGAA	CCGGCCGATC	AAAACGCCTT	GTCCGAACCG	GATGCTGCGA
251	CAGAGGCAGA	GCACTCGGAT	GCGGAAAAAG	CTGCCGACAA	GCAGCCCGTT
301	GCCGATAAAG	CCGACGAGGT	TGAAGAAAAG	GCGGGCGAGC	CGGAACGGGA
351	AGAGCCGGAC	GGACAGGCAG	TGCGTAAGAA	AGCGCTGACG	GAAGAGCGTG
401	AACAAACCGT	CAGGGAAAAA	GCGCAGAAGA	AAGATGCCGA	AACGGTTAAA
451	AAACAAGCGG	TAAACCGTTC	TAAAGAAACA	GAGAAAAAAG	CTTCAAAAGA
501	AGAGAAAAAG	GCGGCGAAGG	AAAAAGTTGC	ACCCAAACCA	ACCCCGGAAC
551	AAATCCTCAA	CAGCGGCAGC	ATCGAAAAAG	CGCGCAGTGC	CGCCGCCAAA
601	GAAGTGACAG	AAATGAAAAA	GTCCGACAAG	GCGGAAGCAA	CGCATTATCT
651	GCAATATGGC	GCGTATGCCG	ACCGTCAGAG	CGCGAAGGG	CAGCGTGCCA



5 This corresponds to the amino acid sequence <SEQ ID 384; ORF65-1>:

	1	MFMNKFQSQG	KGLSGGFFGL	<u>ILATVIIAGI</u>	LFYLNQSQGN	AFKIPASSKQ
	51	PAETELPKK	AQPKEDIQPE	PADQNALSEP	DAATEAQSD	AEKAADKPVP
	101	ADKADVEEK	NQPEREEP	GQAVRKALT	EEREQTVREK	AQKKDAETVK
	151	KQAVKPSKET	EKKASKEEKK	AAKEKVAPKP	TPEQILNSGS	IEKARSAAAK
10	201	EVQMKMTSDK	AMATHYLQMG	AYADRQSAEG	QRAKLAILGI	SSKVVGYQAG
	251	HKTLYRVQSG	NMSADAVKMK	QDELKKHEVA	SLIRSEIS*	

Homology with a predicted ORF from *N.meningitidis* (strain A)

15 *meningitidis:*

[illegible]

40	1	ATGTTTATGA	ACAAATTTTC	CCAATCCGGA	AAAGGTCTGT	CCGTTTTTTT
	51	CTTCGGTTTG	ATACTGGCGA	CGGTTCATTAT	TGCCCGTATT	TTGTTTTATC
	101	TGAACCCAGAG	CGGTCAAAT	GGCTTCAAAA	TCCCGGTTCC	GTCGAAGCAG
	151	CCTGCAGAAA	CGGAAATCCT	GAAACCGGAA	AACCCAGCTA	AGGAAGACAT
	201	CCAACCTGAA	CCGGCCGATC	AAAACGCCTT	GTCCGAACCG	GATGCTGCGA
45	251	AAGAGGCAGA	GCAGTCGGAT	GCGGAAAAAG	CTGCCGACAA	CGACGCCGTT
	301	GCCGACAAAG	CCGACGAGGT	TGAGGAAAAG	CGCGACGAGC	CGGAGCGGGA
	351	AAAGTCGGAC	GGACAGGCAG	TGCGCAAGAA	AGCACTGACG	GAAGAGCGTG
	401	AACAAACCGT	CGGGGAAAAA	GCGCAGAAGA	AAGATGCCGA	AACGGTTAAA
	451	AAACAAGCGG	TAAACCATC	TAAAGAAAAA	GAGAAAAAAG	CTTCAAAGA
50	501	AGAGAAAAAG	CGGGAGAAGG	AAAAAGTTGC	ACCCAAACCG	ACCCCGGAAC
	551	AAATCCTCAA	CAGCGGCAGC	ATCGAAAAAG	CGCGCAGTGC	CGCTGCCAAA
	601	GAAGTGACA	AAATGAAAC	GCCCGACAAG	GCGGAAGCAA	CGCATTTACT
	651	GCAATGGGC	CGGTATGCCG	ACCGCCGGAG	CGCGGAAGGG	CAGCGTGCCA
	701	AACTGGCAAT	TCTGGGCATA	TCTTCCAAAG	TGGTCGGTTA	TCAGGCGGGA
55	751	CATAAAACGC	TTTACCGGGT	GCAAAAGCGGC	AATATGTCTG	CCGATGCGGT
	801	GAAAAAATCG	CAGGACGAGT	TGAAAAACA	TGAAGTCGCC	AGCCTGATCC
	851	GTTCTATCGA	AAGCAAATAA			

This encodes a protein having amino acid sequence <SEQ ID 386>:

1 MFMNKFSSQSG KGLSGFFFL ILATVIIAGI LFYLNQSGQN AFKIPVPSKQ  
 51 PAETEILKPK NQPKEDIQPE PADQNALSEP DAAKEAEQSD AEKAADKQPV  
 101 ADKADEVEEK ADEPEREKSD GQAVRKKALT EEREQTVGEK AQKKAETVK  
 151 KQAVKPSKET EKKASKEEKK AEKEKVAPKP TPEQILNSGS IEKARSAAK  
 201 EVQKMKTPDK AEATHYLQMG AYADRRSAEG QRAKLAILGI SSKVVGQAG  
 251 HKTLYRVQSG NMSADAVKKM QDELKKHEVA SLIRSIESK\*

ORF65a and ORF65-1 show 96.5% identity in 289 aa overlap:

		10	20	30	40	50	60
10	orf65a.pep	MFMNKFSSQSGKGLSGFFFLILATVIIAGILFYLNQSGQN	AFKIPVPSKQPAETEILKPK				
	orf65-1	MFMNKFSSQSGKGLSGFFFLILATVIIAGILFYLNQSGQN	AFKIPASSKQPAETEILKPK				
		10	20	30	40	50	60
15	orf65a.pep	NQPKEDIQPEPADQNALSEPDAAKEAEQSDAEKAADKQPVADKADEVEEKADPEREKSD					
	orf65-1	NQPKEDIQPEPADQNALSEPDAATEAEQSDAEKAADKQPVADKADEVEEKAGEPEREED					
		70	80	90	100	110	120
20	orf65a.pep	GQAVRKKALTEEREQTVGEKAQKKAETVKKQAVKPSKETEKASKEEKKAEKEKVAPKP					
	orf65-1	GQAVRKKALTEEREQTVREKAQKKAETVKKQAVKPSKETEKASKEEKKAAKEKVAPKP					
		130	140	150	160	170	180
25	orf65a.pep	TPEQILNSGSIEKARSAAAKEVQKMKTPDKAEATHYLQMGAYADRRSAEQRAKLAILGI					
	orf65-1	TPEQILNSGSIEKARSAAAKEVQKMKTSKAEATHYLQMGAYADRQSAEQRAKLAILGI					
		190	200	210	220	230	240
30	orf65a.pep	SSKVVGQAGHKTLYRVQSGNMSADAVKKMQDELKKHEVASLIRSIESKX					
	orf65-1	SSKVVGQAGHKTLYRVQSGNMSADAVKKMQDELKKHEVASLIRSIESKX					
		250	260	270	280	290	
35	orf65a.pep	SSKVVGQAGHKTLYRVQSGNMSADAVKKMQDELKKHEVASLIRSIESKX					
	orf65-1	SSKVVGQAGHKTLYRVQSGNMSADAVKKMQDELKKHEVASLIRSIESKX					
		250	260	270	280	290	

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF65 shows 89.6% identity over a 212aa overlap with a predicted ORF (ORF65.ng) from *N.*

gonorrhoeae:

		30	40	50	60	70	80
40	ORF65ng	IIAGILLYLNQGGQNAFKIPAPSKQPAETEILKLNQPKEDIQPEPADQNALSEP	PDVAKE				
	ORF65				ILKPHNQLKEDIQPD	ADQNALSEPDAATE	
45					10	20	30
		90	100	110	120	130	140
50	ORF65ng	AEQSDAEKAADKQPVADKADEVEEKAGEPEREEDGQAVRKKALTEEREQTVREKAQK	QK				
	ORF65	AEQSDAENAADKQPVADKADEVEEKAGEPEREEDGQAVRKKALTEEREQTVREKAQK	QK				
		40	50	60	70	80	90
55	ORF65ng	AETVKKKAVKPSKETEKASKEEKKAAKEKVAPKPTPEQILNSRSIEKARSAAAKEVQ	KM				
	ORF65	AETVKIQAVKPSKETEKASKEEKKAAKEKVAPKPTPEQILNSGSIEKARSAAAKEVQ	KM				
		100	110	120	130	140	150
60	ORF65ng	KNFGQGGSQRIICKWARMPNPGARKGSPVNWQSWAYLPKWSAIRRDIKRFTACKAAIC	PP				
	ORF65	XNVRQGGSXRIICKWARMPNPGARKGSPVNWQSWAYLPKWSAIRRDIKRFTGCKAAIC	LP				
		160	170	180	190	200	210

ORF65ng MR  
 II  
 ORF65 MR

An ORF65ng nucleotide sequence <SEQ ID 387> was predicted to encode a protein having amino

5 acid sequence <SEQ ID 388>:

1 MFMNKFSQSG KGLSGFFFGL ILATVIIAGI LLYLNQGGQN AFKIPAPSKQ  
 51 PAETEILKLK NQPKEDIQPE PADQNALSEP DVAKEAEQSD AEKAADKQPV  
 101 ADKADEVEEK AGEPEREEPD GQAVRKKALT EEREQTVREK AQKKDAETVK  
 151 KKAVKPSKET EKKASKEEKK AAKEKVAPKP TPEQILNSRS IEKARSAAAK  
 201 EVQKMKNFGQ GGSQRIICKW ARMPNPGARK GSVPNWQSWA YLPKWSAIRR  
 251 DIKRFTACKA AICPPMR\*

After further analysis, the complete gonococcal DNA sequence <SEQ ID 389> was found to be:

1 ATGTTTATGA ACAAATTTTC CCAATCCGGA AAAGGTCTGT CCGGTTTCTT  
 15 51 CTTCGGTTTG ATACTGGCAA CGGTTCATTAT TGCCGGTATT TTGCTTTATC  
 101 TGAACCAAGG CGGTCAAATAT GCGTTCAAAA TCCCGGCTCC GTCGAAGCAG  
 151 CCTGCAGAAA CGGAAATCCT GAAACTGAAA AACCAGCCTA AGGAAGACAT  
 201 CCAACCTGAA CCGGCCGATC AAAACGCCTT GTCCGAACCG GATGTTGCGA  
 251 AAGAGGCAGA GCAGTCGGAT GCGGAAAAAG CTGCCGACAA GCAGCCCCGT  
 301 GCCGACAAAG ccgacgAGGT TGAAGAAAAG GcGGgcgAgc cggaACGGa  
 20 351 aGAGCCGGAC ggACAGGCAG TCGCAAGAA AGCACTGAcg gAAGAgcGTG  
 401 AACAAACcgt cagggAAAAA GCGCagaaga AAGATGCCGA AACGgTTAAA  
 451 AAacaaGCg tAaaaccgtc tAAAGAAACa gagaaaaaag cTtcaaaaga  
 501 agagaaaaaag gcgcgcaaaag aaaAAGttgc acccaaacgc accccggaC  
 25 551 aaatcctcaa cagccgCagc atcgaaaaaag cgcgtagtgc cgctgccaaA  
 601 gaAgtgcaGA AAatgaaaaa ctTtgggcaa ggcgGaagcc aacgcattaT  
 651 CTGcaaatgg gcgcgatgc cgaccgtccg gagcgcggaA gggcagcgtg  
 701 ccaaAActggc aAtcttgGgc atatctTccg aagtgtcgG CTATCAGGCG  
 751 GGACATAAAA CGTTTACCG CGTGCAaagc GGCAatatgt ccgccgatgc  
 801 gGTGAAAAAA ATGCAGGACG AGTTGAAAAA GCATGGGGTt gcCAGCCTGA  
 30 851 TCCGTGcgAT TGAAGGCCAA TAA

This encodes the following amino acid sequence <SEQ ID 390>:

1 MFMNKFSQSG KGLSGFFFGL ILATVIIAGI LLYLNQGGQN AFKIPAPSKQ  
 51 PAETEILKLK NQPKEDIQPE PADQNALSEP DVAKEAEQSD AEKAADKQPV  
 101 ADKADEVEEK AGEPEREEPD GQAVRKKALT EEREQTVREK AQKKDAETVK  
 151 KQAVKPSKET EKKASKEEKK AAKEKVAPKP TPEQILNSRS IEKARSAAAK  
 201 EVQKMKNFGQ GGSQRIICKW ARMPTVRS AE GQRAKLAILG ISSEVVGYQA  
 251 GHKTLYRVQS GNMSADAVKK MQDELKKHGV ASLIRAIEGK \*

ORF65ng-1 and ORF65-1 show 89.0% identity in 290 aa overlap:

40 orf65-1.pep MFMNKFSQSGKGLSGFFFGLILATVIIAGILFYLNQSGQNAFKIPASSKQPAETEILKPK  
 orf65ng-1 MFMNKFSQSGKGLSGFFFGLILATVIIAGILLYLNQGGQNAFKIPAPSKQPAETEILKLK  
 45 orf65-1.pep NQPKEDIQPEPADQNALSEPDAAATEAEQSDAEKAADKQPVADKADEVEEKAGEPEREEPD  
 orf65ng-1 NQPKEDIQPEPADQNALSEPDVAKEAEQSDAEKAADKQPVADKADEVEEKAGEPEREEPD  
 50 orf65-1.pep GQAVRKKALTEEREQTVREKAQKKDAETVKKQAVKPSKETEKKASKEEKKAAKEKVAPKP  
 orf65ng-1 GQAVRKKALTEEREQTVREKAQKKDAETVKKQAVKPSKETEKKASKEEKKAAKEKVAPKP  
 55 orf65-1.pep TPEQILNSGSIEKARSAAAKEVQKMKTSDAEATHYL-QMGAYADRS AEGQRAKLAILG  
 orf65ng-1 TPEQILNSRSIEKARSAAAKEVQKMKNFGQGSQRIICKWARMPTVRS AEGQRAKLAILG  
 60

		190	200	210	220	230	240
	240	250	260	270	280	290	
5	orf65-1.pep	ISSKVVGYQAGHKTLYRVQSGNMSADAVKKMQDELKKHEVASLIRSIESKX					
		:					
	orf65ng-1	ISSEVVGYQAGHKTLYRVQSGNMSADAVKKMQDELKKHGVASLIRAIEGKX					
		250	260	270	280	290	

On this basis, including the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 46

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 391>:

15	1	ATGAACCACG	ACATCACTTT	CCTCACCCCTG	TTCCTACTCG	GtKtCTTCGG
	51	CGGAACGCAC	TGCATCGGTA	TGTGCGGCGG	ATTAAGCAGC	GcGTTTGs.s
	101	TCCAACCTCCC	CCCGCATATC	AACCGCTTTT	GGCTGATCCT	GCTGCTTAAC
	151	ACAGGACGGG	TAAGCAGCTA	TACGGCAATC	GGCCTGATAC	TCGGATTAAT
20	201	CGGACAGGTC	GGCGTTTCAC	TCGACCAaAC	CCGCGTCCTG	CAGAATATTT
	251	TATACACGGC	CGCCAACCTC	CTGCTGCTCT	TTTTAGGCTT	ATACTTGAGC
	301	GGTATTTCTT	CCTTGGCGGC	AAAAATCGAG	AAaATCGGCA	AACCGATATG
	351	GCGGAACCTG	AACCCGATAC	TCAACCGGCT	GTTACCCATA	AAATCCATAC
	401	CCGCCTGCCT	tGCGgTCGGA	ATATTATGGG	GCTGGCTGCC	GTGCGGACTG
25	451	GTTTACAGCG	CGTCGCTTTA	CGCGCTGGGA	AgCGGTAGTG	CGGCAACGGG
	501	CGGGTTATAT	ATGCTTGCCT	TTGCACTGGG	TACGCTGCCC	AATCTTtTAG
	551	CAATCGGCAT	TTTTtCCCTG	CAACTGAaWA	AAATCATGCA	AAACCGATAT
	601	ATCCGCCTGT	GTACGGGATT	ATCCGTATCA	TTATGGGCAT	TATGGAAACT
	651	TGCCGTCCTG	TGGCTGTAA			

This corresponds to the amino acid sequence <SEQ ID 392; ORF103>:

30	1	MNHDITFLTL	FLLGXFGGTH	CIGMCGGLSS	AFXQLPPhi	NRFWLILLN
	51	TGRVSSYTAI	GLILGLIGQV	GVSLDQTRVL	QNILYTAANL	LLLFLGLYLS
	101	GISSLAAKIE	KIGKPIWRNL	NPILNRLLPi	KSIPACLAVG	ILWGWLPcGL
	151	VYSASLYALG	SGSAATGGLY	MLAFALGTLP	NLLAIGIFSL	QLKKIMQNRy
	201	IRLCTGLSVS	LWALWKLAVL	WL*		

Further work elaborated the DNA sequence <SEQ ID 393> as:

40	1	ATGAACCACG	ACATCACTTT	CCTCACCCCTG	TTCCTACTCG	GTTTCTTCGG
	51	CGGAACGCAC	TGCATCGGTA	TGTGCGGCGG	ATTAAGCAGC	GCGTTTGCGC
	101	TCCAACCTCCC	CCCGCATATC	AACCGCTTTT	GGCTGATCCT	GCTGCTTAAC
	151	ACAGGACGGG	TAAGCAGCTA	TACGGCAATC	GGCCTGATAC	TCGGATTAAT
	201	CGGACAGGTC	GGCGTTTCAC	TCGACCAaAC	CCGCGTCCTG	CAGAATATTT
	251	TATACACGGC	CGCCAACCTC	CTGCTGCTCT	TTTTAGGCTT	ATACTTGAGC
	301	GGTATTTCTT	CCTTGGCGGC	AAAAATCGAG	AAAATCGGCA	AACCGATATG
	351	GCGGAACCTG	AACCCGATAC	TCAACCGGCT	GTTACCCATA	AAATCCATAC
45	401	CCGCCTGCCT	TGCGGTcGGA	ATATTATGGG	GCTGGCTGCC	GTGCGGACTG
	451	GTTTACAGCG	CGTCGCTTTA	CGCGCTGGGA	AGCGGTAGTG	CGGCAACGGG
	501	CGGGTTATAT	ATGCTTGCCT	TTGCACTGGG	TACGCTGCCC	AATCTTTTAG
	551	CAATCGGCAT	TTTTTCCCTG	CAACTGAAAA	AAATCATGCA	AAACCGATAT
	601	ATCCGCCTGT	GTACGGGATT	ATCCGTATCA	TTATGGGCAT	TATGGAAACT
	651	TGCCGTCCTG	TGGCTGTAA			

This corresponds to the amino acid sequence <SEQ ID 394; ORF103-1>:

1	MNHDITFLTL	FLLGFFGGTH	CIGMCGGLSS	AFALQLPPhi	NRFWLILLN
51	TGRVSSYTAI	GLILGLIGQV	GVSLDQTRVL	QNILYTAANL	LLLFLGLYLS
101	GISSLAAKIE	KIGKPIWRNL	NPILNRLLPi	KSIPACLAVG	ILWGWLPcGL
151	VYSASLYALG	SGSAATGGLY	MLAFALGTLP	NLLAIGIFSL	QLKKIMQNRy

201 IRLCTGLSVS LWALWKLAVL WL\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF103 shows 93.8% identity over a 222aa overlap with an ORF (ORF103a) from strain A of *N.*

5 *meningitidis*:

		10	20	30	40	50	60
orff103.p		MNHDITFLTLFLLGXF	GGTHCIGMCGGLSSAF	XXQLPPHINR	FWLILLNTGRVSS	YTAI	
orff103a		MNXDITFLTLFLLGFF	GGTHCIGMCGGLSSAF	ALQLPPHINRXW	LILLNTGRVSS	YTAI	
	10	20	30	40	50	60	
		70	80	90	100	110	120
orff103.p		GLILGLIGQVGVSLDQ	TRVLQNILYTAANLL	LLFLGLYLSGISSLA	AKIEKIGKPIWRNL		
orff103a		GLILGLIGQVGVSLDQ	TRVXQNILYTAANLL	LLFLGLYLSGISSLA	AKIEKIGKPIWRNL		
	15	70	80	90	100	110	120
		130	140	150	160	170	180
orff103.p		NPILNRLLPIKSIPAC	LAVGILWGWPCGLV	SASLYALGSGSAAT	GGLYMLAFALGTLP		
orff103a		NPILNRLLPIKSIPAC	LAVGILWGWPCGLV	SASLYALGSGSAAT	GGLYMLAFALGTLP		
	20	130	140	150	160	170	180
		190	200	210	220		
orff103.p		NLLAIGIFSLQLXKIM	QNRRI	RLCTGLSVSLWALW	KLAVLWLX		
orff103a		NLXAIGIFSLQLXKIM	QNRRI	RLCTGLSVSLWALW	KLAVLWLX		
	25	190	200	210	220		

The complete length ORF103a nucleotide sequence <SEQ ID 395> is:

30	1	ATGAACCANG	ACATCACTTT	CCTCACCTG	TTCCTACTCG	GTTTCTTCGG
	51	CGGAACGCAC	TGCATCGGTA	TGTGCGGCGG	ATTAAGCAGC	GCGTTTGCGC
	101	TCCAACCTCC	CCCGCATATC	AACCGCTTNT	GGCTGATCCT	GCTGCTTAAC
	151	ACAGGACGGG	TAAGCAGCTA	TACGGCAATC	GGCCTGATAC	TCGGATTAAT
35	201	CGGACAGGTC	GGCGTTTCAC	TCGACCAAAC	CCGCGTCNTG	CAGAATATTT
	251	TATACACGGC	CGCCAACCTC	CTGCTGCTCT	TTTAGGCTT	ATACTTGAGC
	301	GGTATTTCTT	CCTTGGCGGC	AAAATCGAG	AAAATCGGCA	AACCGATATG
	351	GCGGAACCTG	AACCCGATAC	TCAACCGGCT	GTTACCCATA	AAATCCATAC
	401	CCGCTGCTG	TGCGGTGCGA	ATATTATGGG	GCTGGCTGCC	GTGCGGACTA
40	451	GTTTACAGCG	CGTCGCTTTA	CGCGCTGGGA	AGCGGTAGTG	CGGCAACGGG
	501	CGGGTTATAT	ATGCTTGCCT	TTGCACTGGG	TACGCTGCCC	AATCTTTNGG
	551	CAATCGGCAT	TTTTTCCCTG	CAACTGNAAA	AAATCATGCA	AAACCGATAT
	601	ATCCGCCTGT	GTACGGGATT	ATCCGTATCA	TTATGGGCAT	TATGGAAACT
	651	TGCCGTCTCT	TGGCTGTAA			

This encodes a protein having amino acid sequence <SEQ ID 396>:

45	1	MNXDITFLTL	FLLGFFGGTH	CIGMCGGLSS	AFALQLPPI	NRXLILLN
	51	TGRVSSYTAI	GLILGLIGQV	GVSLDQTRVX	QNILYTAANL	LLLFLGLYLS
	101	GISSLAAKIE	KIGKPIWRNL	NPILNRLLPI	KSIPACLAVG	ILWGWPCGL
	151	VYSASLYALG	SGSAATGGLY	MLAFALGTLP	NLXAIGIFSL	QLXKIMQNR
	201	IRLCTGLSVS	LWALWKLAVL	WL*		

50 ORF103a and ORF103-1 show 97.7% identity in 222 aa overlap:

		10	20	30	40	50	60
orff103a.p		MNXDITFLTLFLLGFF	GGTHCIGMCGGLSSAF	ALQLPPHINRXW	LILLNTGRVSS	YTAI	
orff103-1		MNHDITFLTLFLLGFF	GGTHCIGMCGGLSSAF	ALQLPPHINR	FWLILLNTGRVSS	YTAI	
	55	10	20	30	40	50	60
		70	80	90	100	110	120

5	orf103a.pep	GLILGLIGQVGVSLDQTRVXQNILYTAANLLLLFLGLYLSGSISSLAAKIEKIGKPIWRNL
	orf103-1	GLILGLIGQVGVSLDQTRVLQNILYTAANLLLLFLGLYLSGSISSLAAKIEKIGKPIWRNL
10	orf103a.pep	NPILNRLPIKSIPACLA VGILWGWLP CGLVYSASLYALGSGSAATGGLYMLAFALGTLP
	orf103-1	NPILNRLPIKSIPACLA VGILWGWLP CGLVYSASLYALGSGSAATGGLYMLAFALGTLP
15	orf103a.pep	NLXAIGIFSLQLXKIMQNR YIRLCTGLSVSLWALWKLAVLWLX
	orf103-1	NLLAIGIFSLQLKKIMQNR YIRLCTGLSVSLWALWKLAVLWLX

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF103 shows 95.5% identity over a 222aa overlap with a predicted ORF (ORF103.ng) from *N.*

### gonorrhoeae:

20	orf103.pep	MNHDITFLTLFLLGXFGGTHCIGMCGGLSSAFXXQLPPHINRFWLILLNTGRVSSYTAI	60
	orf103ng	MNHDITFLTLFLLGFFGGTHCIGMCGGLSSAFALQLPPHINRFWLILLNTGRISYTAI	60
25	orf103.pep	GLILGLIGQVGVSLDQTRVLQNILYTAANLLLLFLGLYLSGSISSLAAKIEKIGKPIWRNL	120
	orf103ng	GLMLGLIGQLGISLDQTRVLQNILYTAANLLLLFLGLYLSGSISSLAAKIEKIGKPIWRNL	120
30	orf103.pep	NPILNRLPIKSIPACLA VGILWGWLP CGLVYSASLYALGSGSAATGGLYMLAFALGTLP	180
	orf103ng	NPILNRLPIKSIPACLA VGILWGWLP CGLVYSASLYALGSGSATTGGLYMLAFALGTLP	180
35	orf103.pep	NLLAIGIFSLQLXKIMQNR YIRLCTGLSVSLWALWKLAVLWL	222
	orf103ng	NLLAIGIFSLQLKKIMQNR YIRLCTGLSVSLWALWKLAVLWL	222

The complete length ORF103ng nucleotide sequence <SEQ ID 397> is:

40	1	ATGAACCACG	ACATCACTTT	CCTCACCTG	TTCCTGCTCG	GTTTCTTCGG
	51	CGGAATCAC	TGCATCGGT	TGTGCGGCG	ATTAAGCAG	CGGTTTGCG
45	101	TCCAACTCC	CCCGCATAT	AACCGCTTT	GGCTGATTCT	GCTGCTTAAC
	151	ACAGACCGA	TAAGCAGTA	TACGGCAAT	GGCCTGATGC	TCGGATTAAT
50	201	CGGACAAC	GGCATTTC	TCGACCAAA	cgcgcTCCTG	CAAAATATTT
	251	tatacacag	ctccaaCCT	CTGCTGCTCT	TTTAGGCTT	ATACTTGAGC
55	301	GGTATTTCT	CCTTGCGGC	AAAAATCGG	AAAATCGGC	AACCGATATG
	351	GCGCAACCT	AACCCGATC	TCAACCGCT	GCTGCCCAT	AAATCCATAC
60	401	CCGCTGCCT	TGCTGTCGA	ATATTATGG	GCTGGCTGCC	GTGCGGACTG
	451	GTTTACAGC	CATCACTTA	CGCGCTGGA	AGCGGTAGT	CGACAACCGG
65	501	CGGACTGTAT	ATGCTTGCT	TTGCACTGG	TACGCTGCC	AATCTTTTGG
	551	CAATCGGCAT	TTTTCCCTG	CAACTGAAA	AAATCATGA	AAACCGATAT
70	601	ATCCGCCTGT	GTACAGGAT	ATCCGTATCA	TTATGGGCAT	TATGGAAGCT
	651	TGCCGCTCTG	TGGCTGTAA			

This encodes a protein having amino acid sequence <SEQ ID 398>:

55	1	MNHDITFLTL	FLLGFFGGTH	CIGMCGGLSS	AFALQLPPI	NREWLILLN
	51	TGRISYTAI	GLMLGLIGQL	GISLDQTRVL	QNILYTASN	LLLLFLGLYLS
60	101	GISSLAAKIE	KIGKPIWRNL	NPILNRLPI	KSIPACLA VG	ILWGWLP CGL
	151	VYSASLYALG	SGSATTGGGLY	MLAFALGTLP	NLLAIGIFSL	QLKKIMQNR Y
65	201	IRLCTGLSVS	LWALWKLAVL	WL*		

In addition, ORF103ng and ORF103-1 show 97.3% identity in 222 aa overlap:

60	orf103-1.pep	MNHDITFLTLFLLGFFGGTHCIGMCGGLSSAFALQLPPHINRFWLILLNTGRVSSYTAI

	orf103ng	MNHDITFLTLFLLGFFGGTHCIGMCGGLSSAFALQLPPHINRFWLILLNLTGRISSTAI	10	20	30	40	50	60
			70	80	90	100	110	120
5	orf103-1.pep	GLILGLIGQVGVSLDQTRVLQNILYTAANLLLLFLGLYLSGISSLAAKIEKIGKPIWRNL						
	orf103ng	GLMLGLIGQLGISLDQTRVLQNILYTAANLLLLFLGLYLSGISSLAAKIEKIGKPIWRNL	70	80	90	100	110	120
10	orf103-1.pep	NPILNRLPIKSIPACLA VGILWGWLP CGLVYSASLYALGSGSAATGGLYMLAFALGTL P	130	140	150	160	170	180
	orf103ng	NPILNRLPIKSIPACLA VGILWGWLP CGLVYSASLYALGSGSAT TGGLYMLAFALGTL P	130	140	150	160	170	180
15	orf103-1.pep	NLLAIGIFSLQLKKIMQNRYIRLCTGLSVSLWALWKLAVLWLX	190	200	210	220		
	orf103ng	NLLAIGIFSLQLKKIMQNRYIRLCTGLSVSLWALWKLAVLWLX	190	200	210	220		

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 47

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 399>:

	1	ATGGAAAACC	AAAGGCCGCT	CCTAGGCTTT	CGCTTGGCAC	TTTGGCGGCG
	51	GATGACGTGG	GGACGCTGC	CGAT.TCCGT	GCGGCAGGTA	TTGAAGTTTG
30	101	TCGATGCGCC	GACGCTGGTG	TGGGTGCGTT	TTACCGTGCG	GGCGGCGGTA
	151	TTGTTTGTTT	TGCTGGCACT	GGCGGGGCGG	CTGCcGAAGC	GGCGaGGATT
	201	TTTCTTGGTG	CTCATT CAGG	CTGCTGCTGC	TCGGCGTGCG	GGGCATTTTCG
	251	GCAAACTTTG	TGCTGATTGC	CCAAGGGCTG	CATTATATTT	CGCCGACCAC
	301	GACGCAGGTT	TTGTGGCAGA	TTTCGCCGTT	TACGATGATT	GTWGTCCGTTG
35	351	TGTTGGTGTT	TAAAGACCGG	ATGACTGCCG	CTCAGAAAAT	CGGCTTGTTT
	401	TTGCTGCTTG	CCGTTTGCT	TATGTATTTT	AACGATAAAT	TCGGCGAGTT
	451	GTCGGGTTTG	GGCGCGTATG	C.AAGGGCGT	GTTGCTGTGT	GCGGCAGGCA
	501	GTATGGCATG	GGTGTGTAAT	GCCGTGGCGC	AAAAGCTGCT	GTGCGCGCAA
	551	TTCGGGCCCG	AACAGATTCT	GCTGTTGATT	TATGCGGCAA	GTGCGGCCGT
40	601	GTTCCTGCCG	TTTGCCGAAC	CGGCACACAT	CGGAAGTATG	GACCGTACGT
	651	TGGCGTGGGT	ATGTATTGCG	TATTGCTGCT	TGAATACGTT	AATCGGTTAC
	701	GGCTCGTTTC	GCGAGGCGTT	GAAACATTGG	GAGGCTTCCA	AAGTCAGCGC
	751	GGTAACAACC	TTGCTCCCCG	TGTTTACCGT	AATAAATACT	TTGCTCGGGC
	801	ATTATGTGAT	GCCTGAAACT	TTTGCCGCGC	CGGA..	

This corresponds to the amino acid sequence <SEQ ID 400; ORF104>:

45	1	MENQRPLLGF	RLALLAAMTW	GTLPXSVRQV	LKFVDAPTLV	WVRFTVAAAV
	51	LFVLLALGGR	LPKRDFSWC	SFRLLLLGVA	GISANFVLIA	QGLHYISPTT
	101	TQVLWQISPF	TMIVVGVLVF	KDRMTAAQKI	GLVLLLAGLL	MYFNDKFGEI
	151	SGLGAYXKGV	LLCAAGSMW	VCNAVAQKLL	SAQFGPQQIL	LLIYAASAAV
50	201	FLPFAEPAHI	GSM DGT LAWV	CIAYCCLNTL	IGYGSFGEAL	KHWEASKVSA
	251	VTILLPVFTV	INTLLGHYVM	PETFAAP...		

Further work revealed further partial DNA sequence <SEQ ID 401>:

	1	ATGGAAAACC	AAAGGCCGCT	CCTAGGCTTC	GCGTTGGCAC	TTTGGCGGCG
	51	GATGACGTGG	GGACGCTGC	CGATTGCCGT	GCGGCAGGTA	TTGAAGTTTG
55	101	TCGATGCGCC	GACGCTGGTG	TGGGTGCGTT	TTACCGTGCG	GGCGGCGGTA
	151	TTGTTTGTTT	TGCTGGCACT	GGCGGGGCGG	CTGCCGAAGC	GGCGGGATTT
	201	TTCTTGGTGC	TCATT CAGGC	TGCTGCTGCT	GCGCGTGCGC	GGCATTTCGG

5  
10  
251 CAAACTTTGT GCTGATTGCC CAAGGGCTGC ATTATATTTT GCCGACCACG  
301 ACGCAGGTTT TGTGGCAGAT TTCGCCGTTT ACGATGATTG TTGTCGGTGT  
351 GTTGGTGTGT AAAGACCGGA TGAAGTCCGC TCAGAAAATC GGCTTGGTTT  
401 TGCTGCTTGC CGGTTTGCTT ATGTTTTTTA ACGATAAATT CGGCGAGTTG  
451 TCGGGTTTGG GCGCGTATGC GAAGGGCGTG TTGCTGTGTG CGGCAGGCAG  
501 TATGGCATGG GTGTGTTATG CCGTGGCGCA AAAGCTGCTG TCGGCACAAT  
551 TCGGGCCGCA ACAGATTCTG CTGTTGATTT ATGCGGCAAG TGCCGCGGTG  
601 TTCCTGCCGT TTGCCGAACC GGCACACATC GGAAGTTTGG ACGGTACGTT  
651 GGCCTGGGTT TGTTTTGCGT ATTGCTGCTT GAATACGTTA ATCGGTTACG  
701 GCTCGTTCGG CGAGGCGTTG AAACATTGGG AGGCTTCCAA AGTCAGCGCG  
751 GTAACAACCT TGCTCCCCGT GTTTACCGTA ATAwTwCCTT TGCTCGGGCA  
801 TTATGTGATG CCTGAAACTT TTGCCGCGCC GGA...

This corresponds to the amino acid sequence <SEQ ID 402; ORF104-1>:

15  
251 MENQRPLLGF ALALLAAMTW GTLPPIAVRQV LKFVDAPTLV WVRFTVAAAV  
51 LFVLLALGGR LPKRRDFSWC SFRLLLLGVA GISANFVLI A QGLHYISPTT  
101 TQVLWQISPF TMIVVGVLVF KDRMTAAQKI GLVLLLAGLL MEFNDKFGEL  
151 SGLGAYAKGV LLCAAGSMAW VCYAVAQKLL SAQFGPQQIL LLIYAASAAV  
201 FLPPFAEPAHI GSLDGTALAV CFAYCCLNTL IGYGSFGEAL KHWEASKVSA  
251 VTLLPVFTV IXXLLGHYVM PETFAAP...

20 Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical HI0878 protein of *H. influenzae* (accession number U32769)

ORF104 and HI0878 show 40% aa identity in 277aa overlap:

25  
30  
35  
40  
orf104 4 QRPLLGFRLLALLAAMTWGTLPSVSRQVLKFVDAPTLVWXXXXXXXXXXXXXXXXXXXXP- 62  
Q+PLLGF AL+ AM WG+LP +++QVL ++A T+VW P  
HI0878 3 QQPLLGFTFALITAMAWGSLPIALKQVLSVMNAQTIVWYRFIIAAVSLALLAYKKQLPE 62  
orf104 63 --KRRDFSWCSFRLLLLGVAGISANFVLI A QGLHYISPTTTQVLWQISPFMTIVVGVLVF 120  
K R ++W ++L+GV G+++NF+L + L+YI P+ Q+ +S F M++ GVL+F  
HI0878 63 LMKVRQYAW----IMLIGVIGLTSNFLLFSSSLNYIEPSVAQIFIHLSFGLICGVLI 118  
orf104 121 KDRMTAAQKIXXXXXXXXXXXMYFNDKFGELSGLGAYXKGVLLCAAGSMAWVCNAVAQKLL 180  
K+++ QKI ++FND+F +GL Y GV+L G++ WV +AQKL+  
HI0878 119 KEKLGHLQKIGLFLLLIGLGFNDRFDAFAGLNQYSTGVILGVGGALIWWAYGMAQKLM 178  
orf104 181 SAQFGPQQIILLIYAASAAVFLPPFAEPAHIGSMDGTALAWVCIAYCCLNTLIGYGSFGEAL 240  
+F QQILL++Y A F+P A+ + + + LA +C YCCLNTLIGYGS+ EAL  
HI0878 179 LRKFNSQQIILLMYLGCAIAFMFMAFDSQVQELT-PLALICFIYCCLNTLIGYGSYAEAL 237  
orf104 241 KHWEASKVSAVTTLLPVFTVINTLLGHYVMPETFAAP 277  
W+ SKVS V TL+P+FT++ + + HY P FAAP  
HI0878 238 NRWDVSKVSVITLVLFTILFSLIAHYFSPADFAAP 274

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF104 shows 95.3% identity over a 277aa overlap with an ORF (ORF104a) from strain A of *N.*

45 *meningitidis*:

50  
55  
orf104.pep 10 20 30 40 50 60  
MENQRPLLGFRLALLAAMTWGTLPSVSRQVLKFVDAPTLVWVRFTVAAAVLFVLLALGGR  
||||| : |||||  
orf104a 10 20 30 40 50 60  
MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFTVAAAVLFVLLALGGR  
70 80 90 100 110 120  
orf104.pep LPKRRDFSWCSFRLLLLGVAGISANFVLI A QGLHYISPTTTQVLWQISPFMTIVVGVLVF  
||| : |||||  
orf104a 70 80 90 100 110 120  
LPKWRDFSWCSFRLLLLGVAGISANFVLI A QGLHYISPTTTQVLWQISPFMTIVVGVLVF  
130 140 150 160 170 180  
orf104.pep KDRMTAAQKIGLVLLLAGLLMYFNDKFGELSGLGAYXKGVLLCAAGSMAWVCNAVAQKLL



5	orf104a	:	KDRMTAAQKIGLVLLLAGLLMFFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL	130	140	150	160	170	180
				190	200	210	220	230	240
10	orf104.pep		SAQFGPQQIILLIYAASAAVFLPFAEPAHIGSMDGTLAWVCIAYCCLNTLIGYGSFGEAL						
	orf104a	:	SAQFGPQQIILLIYAASAAVFLPFAELAHIGSLDGTSLAWVCFAYCCLNTLIGYGSFGEAL	190	200	210	220	230	240
15	orf104.pep		:	250	260	270			
	orf104a	:	KHWEASKVSAVTTLLPVFTVIFSLLGHYVMPDTFAAPDMNGLYAGALVVVGAVTAAVG	250	260	270	280	290	300

The complete length ORF104a nucleotide sequence <SEQ ID 403> is:

1	ATGGAACC	AAAGCCGCT	CCTAGGCTTC	GCGTTGGCAC	TTTGGCGGC
51	GATGACGTGG	GGAACGCTGC	CGATTGCCGT	GCGGCAGGTA	TTGAAGTTTG
101	TCGATGCGCC	GACGCTGGTG	TGGGTGCGTT	TTACCGTGGC	GGCGGCGGTA
20	151	TTGTTTGT	TGCTGGCATT	GGGCGGCGG	CTGCCGAAGT
	201	TTCTTG	TCATTCAGGC	TGCTGCTGCT	CGGCGTGCG
	251	CAAACTTTGT	GCTGATTGCC	CAAGGGCTGC	ATTATATTTC
	301	ACGCAGGTTT	TGTGGCAGAT	TTCGCCGTTT	ACGATGATTG
	351	GTTGGTGT	AAAGACCGGA	TGACTGCCGC	TCAGAAAATC
25	401	TGCTGCTTGC	CGGTTTGCTT	ATGTTTTTTA	ACGATAAATT
	451	TCGGGTTTGG	GCGCGTATGC	GAAGGGCGTG	TTGCTGTGTG
	501	TATGGCATGG	GTGTGTTATG	CCGTGGCGCA	AAAGCTGCTG
	551	TCGGGCCGCA	ACAGATTCTG	CTGTTGATTT	ATGCGGCAAG
	601	TTCTGCCGT	TTGCCGAAC	GGCACACATC	GGAAGTTTGG
30	651	GGCGTGGGTT	TGTTTTGCGT	ATTGCTGCTT	GAATACGTTA
	701	GCTCGTTCGG	CGAGGCGTTG	AAACATTGGG	AGGCTTCCAA
	751	GTAACAACCT	TGCTCCCGT	GTTTACCGTA	ATATTTTCTT
	801	TTATGTGATG	CCTGATACTT	TTGCCGCGCC	GGATATGAAC
35	851	ATGCCGCGC	ACTGGTCTGT	GTCGGGGGTG	CGGTTACGCG
	901	GACAGGCTGT	TCAAACGCCG	CTAG	

This encodes a protein having amino acid sequence <SEQ ID 404>:

1	MENQRPL	GFALALLA	AMTWGTL	PIAVRQV	LKFVDAPT	LVWVRFT	VAAAVL
51	LFVLLAL	GGRLPKWR	DFSWCS	SFRLLLL	GVA	GISANFV	LIAQGL
101	TQVLWQ	ISPTF	TMIVVGV	LVFKDR	MTAAQKI	GLVLLAG	LLMFFN
40	151	SLGAYAK	GVLLCAAG	SMAWVCY	AVAQKLL	SAQFGPQ	QILLLI
	201	FLPFAEL	AHIGSLD	GTSLAWV	CFAYCCL	NLTI	IGYGSF
	251	VTTLLPV	FTVIFSL	LGHVMP	DTFAAP	DMNGLY	AGALVV
	301	DRLF	KRR*				

ORF104a and ORF104-1 show 98.2% identity in 277 aa overlap:

45	orf104a.pep	10	20	30	40	50	60
	orf104-1	10	20	30	40	50	60
50	orf104a.pep	70	80	90	100	110	120
	orf104-1	70	80	90	100	110	120
55	orf104a.pep	130	140	150	160	170	180
	orf104-1	130	140	150	160	170	180
60	orf104a.pep	190	200	210	220	230	240
	orf104-1	190	200	210	220	230	240

5	orf104a.pep	SAQFGPQQILLIYAASAAVFLPFAELAHIGSLDGT LAWVCFAYCCLNTLIGYGSFGEAL				
	orf104-1	SAQFGPQQILLIYAASAAVFLPFAEPAHIGSLDGT LAWVCFAYCCLNTLIGYGSFGEAL	190	200	210	220 230 240
10	orf104a.pep	KHWEASKVSAVTTLLPVFTVIFSL LGHYVMPDTFAAPDMNGLGYAGALVVVGAVTAAVG	250	260	270	280 290 300
	orf104-1	KHWEASKVSAVTTLLPVFTVIXLLGHYVMPETFAAP	250	260	270	

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF104 shows 93.9% identity over a 277aa overlap with a predicted ORF (ORF104.ng) from *N.gonorrhoeae*:

15	orf104.pep	MENQRPLLGFRLALLAAMTWGTL PXSVRQVLKFVDAPTLVWVRFTVAAAVLFVLLALGGR	60
	orf104ng	MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFTVAAAVLFVLLALGGR	60
20	orf104.pep	LPKRRDFSWSFRLLLLGVAGISANFVLIAQGLHYISPTTTQVLWQISPTTMI VVGVLVF	120
	orf104ng	LPKRRDFSWSFRLLLLGVGTGISANFVLIAQGLHYISPTTTQVLWQISPTTMI VVGVLVF	120
25	orf104.pep	KDRMTAAQKIGLVLLLAGLLMYFNDKFGE SGLGAYXKGVLLCAAGSMAWVCNAVAQKLL	180
	orf104ng	KDRMTAAQKIGLVLLVGLLMFFNDKFGE SGLGAYAKGVLLCAAGSMAWVCYAVAQKLL	180
30	orf104.pep	SAQFGPQQILLIYAASAAVFLPFAEPAHIGSMDGT LAWVCIAYCCLNTLIGYGSFGEAL	240
	orf104ng	SAQFGPQQILLIYAASAAVFLXAEPAHIGSLDGT LAWVCFVYCCLNTLIGYGSFGEAL	240
35	orf104.pep	KHWEASKVSAVTTLLPVFTVINTLLGHYVMPETFAAP	277
	orf104ng	KHWEASKVSAVTTLLPVFTVIFSL LGHYVMPDTFAAPDMNGLGYVGALVVVGAVTAAVG	300

The complete length ORF104ng nucleotide sequence <SEQ ID 405> is predicted to encode a protein having amino acid sequence <SEQ ID 406>:

40	1	MENQRPLLGF ALALLAAMTW	GTLP	IAVRQV	LKFVDAPTLV	WVRFTVAAAV
	51	LFVLLALGGR	LPKRRDFS	SWH	SFRLLLLGVT	GISANFVLIA QGLHYISPTT
45	101	TQVLWQISPF	TMIVVGVLVF	KDRMTAAQKI	GLVLLLVGLL	MEFNDKFGE L
	151	SGLGAYAKGV	LLCAAGSMAW	VCYAVAQKLL	SAQFGPQQIL	LLIYAASAAV
50	201	FLLXAEPAH	IGSLDGT LAWV	CFVYCCLNTL	IGYGSFGEAL	KHWEASKVSA
	251	VTTLLPVFTV	IFSL LGHYVM	PDTFAAPDMN	GLGYVGALVV	VGGAVTAAVG
	301	DRPFKRR*				

Further work revealed the complete gonococcal nucleotide sequence <SEQ ID 407>:

45	1	ATGGAAAACC	AAAGGCCGCT	CCTAGGCTTC	GCGTTGGCAC	TTTGGCGGC
	51	GATGACGTGG	GGGACGCTGC	CGATTGCCGT	GCGGCAGGTA	TTGAAGTTTG
50	101	TCGATGCGCC	GACGCTGGTG	TGGGTGCGTT	TTACCGTGGC	GGCGGCGGTA
	151	TTGTTTGTTT	TGCTGGCATT	GGGCGGGCGG	CTGCCGAAGC	GGCGGGATTT
55	201	TTCTTGGCAT	TCATT CAGGC	TGCTGCTGCT	CGGCGTGACG	GGCATTTCGG
	251	CAAACTTTGT	GCTGATTGCC	CAAGGGCTGC	ATTATATTTC	GCCGACCACG
60	301	ACGCAGGTTT	TGTGGCAGAT	TTGCGCGTTT	ACGATGATTG	TTGTCGCGGT
	351	GTTGGTGTTT	AAAGACCGGA	tgaCTGCCGC	GCAGAAAATC	GGTTTGTTT
65	401	TGCTGCTtgT	CGGTttgCTT	ATGTTTTtta	ACGACAAAT	CGGCGAGTTG
	451	TCGGGTTTGG	GCGCGTATGC	GAAGGGCGTG	TTGCTGTGTG	CGGCAGGCAG
70	501	TATGGCCTGG	GTGTGTTATG	CCGTGGCGCA	AAAGCTGCTG	TCGGCGCAAT
	551	TCGGGCCGCA	ACAGATTCTG	CTGTTGATTT	ATGCGGcaag	tgccgccGTG
75	601	TTCCtgccgT	TTGccgaacc	GGCACACATC	GGAAGTTTgg	aCGGTACgtt
	651	GGCGTGGGTT	TGTTTTGTGT	ATTGCTGCTT	GAATACGTTA	ATCGGTTACG
80	701	GCTCGTTCGG	CGAGGCGTTG	AAACATTGGG	AGGCTTCCAA	AGTCAGCGCG
	751	GTAACAACCT	TGCTCCCCGT	GTTTACCGTA	ATATTTTCTT	TGCTCGGGCA
	801	TTATGTGATG	CCTGATACTT	TTGCCGCGCC	GGATATGAAC	GGTTTGGGTT

851 ATGTCGGCGC ACTGGTCGTG GTCGGGGGTG CGGTTACGGC GCGCGTGGGG  
901 GACAGGCCGT TCAAACGCCG CTAG

This corresponds to the amino acid sequence <SEQ ID 408; ORF104ng-1>:

```

1  MENQRPLLGF ALALLAAMTW GTLPPIAVRQV LKFVDAPTLV WVRFTVAAAV
5  51  LFVLLALGGR LPKRRDFSWH SFRLLLLGVT GISANFVLI QGLHYISPTT
    101  TQVLWQISPF TMIVVGVLVF KDRMTAAQKI GLVLLLVGLL MEFNDKFGEL
    151  SGLGAYAKG LLCAAGSMAW VCYAVAQKLL SAQFGPQQIL LLIYAASAAV
    201  FLPPFAEPAH GSLDGT LAWV CFVYCCLNTL IGYGSFGEAL KHWEASKVSA
    251  VTTLLPVFTV IFSLLGHYVM PDTFAAPDMN GLGYVGALVV VGGAVTAAVG
10  301  DRPFKRR*

```

ORF104ng-1 and ORF104-1 show 97.5% identity in 277 aa overlap:

```

15  orf104-1.pep  10      20      30      40      50      60
    orf104ng-1  MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFTVAAAVLFVLLALGGR
    10      20      30      40      50      60
20  orf104-1.pep  70      80      90      100     110     120
    orf104ng-1  LPKRRDFSWCSFRLLLLGVAGISANFVLIQGLHYISPTTTQVLWQISPFMTIVVGVLVF
    70      80      90      100     110     120
25  orf104-1.pep  130     140     150     160     170     180
    orf104ng-1  KDRMTAAQKIGLVLLLAGLMMFFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL
    130     140     150     160     170     180
30  orf104-1.pep  190     200     210     220     230     240
    orf104ng-1  SAQFGPQQIILLIYAASAAVFLPPFAEPAHIGSLDGT LAWVCFAYCCLNTLIGYGSFGEAL
    190     200     210     220     230     240
35  orf104-1.pep  250     260     270
    orf104ng-1  KHWEASKVSAVTTLLPVFTVIXXLLGHYVMPETFAAP
    250     260     270
40  orf104ng-1  KHWEASKVSAVTTLLPVFTVIFSLLGHYVMPDTFAAPDMNGLGYVGALVVVGGAVTAAVG
    250     260     270     280     290     300

```

In addition, ORF104ng-1 shows significant homology with a hypothetical *H. influenzae* protein:

gi|1573895 (U32769) hypothetical [Haemophilus influenzae] Length = 306

Score = 237 bits (598), Expect = 8e-62

Identities = 114/280 (40%), Positives = 168/280 (59%), Gaps = 8/280 (2%)

```

45  Query: 30  QRPXXXXXXXXXXMTWGTLPPIAVRQVLKFVDAPTLVWXXXXXXXXXXXXXXXXXXXXP- 88
    Q+P          M WG+LPIA++QVL  ++A T+VW          P
    Sbjct: 3  QQPLLGFTFALITAMAGSLPIALKQVLSVMNAQTIVWYRFIIAAVSLALLAYKKQLPE 62
50  Query: 89  --KRRDFSWHSFRLLLLGVTGISANFVLIQGLHYISPTTTQVLWQISPFMTIVVGVLVF 146
    K R ++W    ++L+GV G+++NF+L +  L+YI P+  Q+  +S F M++ GVL+F
    Sbjct: 63  LMKVRQYAW----IMLIGVIGLTSNFFLSSSLNYIEPSVAQIFIHLSSFGMLICGVLI 118
55  Query: 147  KDRMTAAQKIXXXXXXXXXXMMFFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL 206
    K+++  QKI          +FFND+F  +GL Y+  GV+L  G++ WV Y +AQKL+
    Sbjct: 119  KEKLGHLQKIGLFLLLIGLGLFFNDRFADFAGLNQYSTGVILGVGALIIWVAYGMAQKLM 178
60  Query: 207  SAQFGPQQIILLIYAASAAVFLPPFAEPAHIGSLDGT LAWVCFVYCCLNTLIGYGSFGEAL 266
    +F  QQILL++Y  A F+P A+ + +  L  LA +CF+YCCCLNTLIGYGS+  EAL
    Sbjct: 179  LRKFNSQQILLMMYLGCAIEMPMADFSQVQELT-PLALICFIYCCCLNTLIGYGSYAEAL 237
    Query: 267  KHWEASKVSAVTTLLPVFTVIFSLLGHYVMPDTFAAPDMN 306
    W+ SKVS V TL+P+FT++FS + HY  P  FAAP++N

```

Sbjct: 238 NRWDVSKVSVVITLVPLFTILFSHIAHYFSPADFAAPELN 277

Based on this analysis, including the presence of a putative leader sequence and several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 48

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 409>:

```
1  ATGGTAGCTC  GTCGGGCTCA  TAACCCGAAG  GTCGTAGGTT  CGAATCCTGT
51  .CCCGCAACC  TAATTTCAAA  CCCCTCGGTT  CAATGCCGAG  GG.GTTTGT
101 T.TTGCTGT  TTCCTGTTTC  CTGTTTCCTG  CCGCCTCCGT  TTTTTCGCGG
151 ATTTTCCTTC  CGGCCGCAAT  ATCGGAACGG  CAGACCGCCG  TCTGTTTTCG
201 GTTGCAAATT  CAGGCAGTTT  GGCTACAATC  TTCCGCATTG  TCTTCAAGAA
251 AGCCAACCAT  GCCGACCGTC  CGTTTTACCG  AATCCGTCAG  CAAACAAGAC
301 CTTGATGCTC  TGTTCGAGTG  GGCAAAAGCA  AGTTACGGTG  CAGAAAGTTG
351 CTGGAACACG  CTGTATCTGA  ACGGTCysCC  TTTGGGCAAC  CTGTCGCCGG
401 AATGGGTGGA  ACGCGTsmmA  AAAGACTGGG  AGGCAGGCTG  CyCGGAGTCT
451 TCAGACGGCA  TTTTCTGAA  TgCGGACGGc  TgGcctGATA  TGGgCGGAcg
501 cTTACAGCAC  CTCGCCCTCG  GTTGGCACTG  TCGGGGGCTG  TTGGACGgsT
551 GCGCAACGA  GTGTTTCGAC  CTGACCGACG  GCGGCGGCAA  CCCCTTGTTC
601 ACGCTCGaAc  GCGCCGyTT  mCGTCCTkTC  GGACTGCTCA  GCCGCGCCGT
651 CCATCTCAAC  GGCTGACCG  AATCGGACGG  CCGATGGCAT  TTTGCGATAG
701 GCAGGCGCAG  TCCGCACAAA  GCAGTCGATC  CCAACAAACT  CGACAATACT
751 rCCGCGGCG  GTGTTTCGCG  CGGCGAAATG  CCGTCTGAAG  CCGTGTGTCG
801 CGAAAGCAGC  GAAGAAGCCG  GTTTGGATAA  AACGCTGcTT  CCGCTCATCC
851 GCCCGGTATC  GCAGCTGCAC  AGCCTGCGCT  CCGTCAGCCG  GGGTGTACAC
901 AATGAAATCC  TGTATGTATT  CGATGCCGTC  CTGCCG...
```

This corresponds to the amino acid sequence <SEQ ID 410; ORF105>:

```
1  MVARRAHNPK  VVGSNXPAT  XFQTPRFNAE  XVLXLPVSCF  LFPAASVFCR
51  IFLPAAISER  QTAVALRLQI  QAVWLQSSAL  SSRKPTMPTV  RFTESVSKQD
101 LDALFEWAKA  SYGAESCWKT  LYLNQXPLGN  LSPEWVERVX  KDWEAGCXES
151 SDGIFLNADG  WPMGGRLQH  LALGWHCAGL  LDGWRNECFD  LTDGGGNPLF
201 TLERAXRXPX  GLLSRAVHLN  GLTESDGRWH  FWIGRRSPHK  AVDPKNLNT
251 XAGGVSGGEM  PSEAVCRESS  EEAGLDKTL  PLIRPVSQIH  SLRSVSRGVH
301 NEILYVFDV  LP...
```

Further work revealed the complete nucleotide sequence <SEQ ID 411>:

```
1  ATGCCGACCG  TCCGTTTTAC  CGAATCCGTC  AGCAAACAAG  ACCTTGATGC
51  TCTGTTTCGAG  TGGGCAAAAG  CAAGTTACGG  TGCAGAAAGT  TGCTGGAAAA
101 CGCTGTATCT  GAACGGTCTG  CCTTTGGGCA  ACCTGTCGCC  GGAATGGGTG
151 GAACGCGTCA  AAAAAGACTG  GGAGGCAGGC  TGCTCGGAGT  CTTAGACGG
201 CATTTTTCTG  AATGCGGACG  GCTGGCCTGA  TATGGGCGGA  CGCTTACAGC
251 ACCTCGCCCT  CGGTTGGCAC  TGTGCGGGGC  TGTGGACGG  CTGGCGCAAC
301 GAGTGTTCG  ACCTGACCGA  CGGCGGCGGC  AACCCTTGT  TCACGCTCGA
351 ACGGCGCGCT  TTCCGTCTCT  TCGGACTGCT  CAGCCGCGCC  GTCCATCTCA
401 ACGGTCTGAC  CGAATCGGAC  GGCAGATGGC  ATTTCTGGAT  AGGCAGGCGC
451 AGTCCGCACA  AAGCAGTCGA  TCCCAACAAA  CTCGACAATA  CTGCCGCGCG
501 CGGTGTTTCC  GCGGCGGAAA  TGCCGTCTGA  AGCCGTGTGT  CGCGAAAGCA
551 GCGAAGAAGC  CGGTTTGGAT  AAAACGCTGC  TTCCGCTCAT  CCGCCCGGTA
601 TCGCAGCTGC  ACAGCCTGCG  CTCGTCAGC  CGGGGTGTAC  ACAATGAAAT
651 CCTGTATGTA  TTGATGCCG  TCCTGCCCGA  AACCTTCCTG  CCTGAAATC
701 AGGATGGCGA  AGTGGCGGGT  TTTGAGAAAA  TGGACATCGG  CGGTCTGTTG
751 GATGCCATGT  TGTGCGGAAA  CATGATGCAC  GACGCGCAAC  TGTTACGCT
801 GGACGCGTTT  TGCCGTTACG  GTCTGATTGA  TGCCGCCCAT  CCGCTGTCCG
851 AGTGGCTGGA  CGGCATACGT  TTATAG
```

This corresponds to the amino acid sequence <SEQ ID 412; ORF105-1>:

5

1	MPTVRFTESV	SKQDLDALFE	WAKASYGAES	CWKTLTYLNL	PLGNLSPEWV
51	ERVKKDWEAG	CSESSDGI FL	NADGWPDMMG	RIQHLAGLWH	CAGLLDQWRN
101	ECFDLTGGG	NPLFTLERRA	FRFPGLLSRA	VHLNGLTESD	GRWHFWITGR
151	SPHKAVDPNK	LDNTAAGGVS	GGEMPSEAVC	RESSEAGLD	FTLLPLIRPV
201	SQHLSLRSV	RGVHNEILYV	FDVLPETFL	PENQDGEVAG	FEKMDIGLL
251	DAMLSGRMMH	DAQVLTLDAF	CRYGLIDAAH	PLSEWLDGIR	L*

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF105 shows 89.4% identity over a 226aa overlap with an ORF (ORF105a) from strain A of *N.*

10 *meningitidis:*

		60	70	80	90	100	110
	orf105.pep	I S E R Q T A V C L R L Q I Q A V W L Q S S A L S S R K P T M P T V R F T E S V S K Q D L D A L F E W A K A S Y G A E S					
					:		
15	orf105a				M P T V R F T E S V S K H D L D A L F E W A K A S Y G A E S		
					10	20	30
		120	130	140	150	160	170
	orf105.pep	C W K T T L Y L N G X P L G N L S P E W V E R V X K D W E A G C X E S S D G I F L N A D G W P D M G G R L Q H L A L G W H					
		:                               :					
20	orf105a	C W K T T L Y L N G L P L G N L S P E W A E R V X K D W E A G C S E S S D G I F L N A D G W P D M G R R L Q H L A R I W K					
		40	50	60	70	80	90
		180	190	200	210	220	230
	orf105.pep	C A G L L D G W R N E C F D L T D G G G N P L F T L E R A X X R P X G L L S R A V H L N G L T E S D G R W H F W I G R R					
		:           :					
25	orf105a	E A G L L H G W R D E C F D L T D G G S N P L F A L E R A A F R P F G L L S R A V H L N G L V E S D G R W H F W I G R R					
		100	110	120	130	140	150
		240	250	260	270	280	290
	orf105.pep	S P H K A V D P N K L D N T X A G G V S G G E M P S E A V C R E S S E E A G L D K T L L P L I R P V S Q L H S L R S V S					
		:                   :					
30	orf105a	S P H K A V D P D K L D N T A A G G V S S G E L P S E T V C R E S S E E A G L D K T L L P L I R P V S Q L H S L R P V S					
		160	170	180	190	200	210
35		300	310				
	orf105.pep	R G V H N E I L Y V F D A V L P					
	orf105a	R G V H N E I L Y V F D A V L P E T F L P E N Q D G E V A G F E K M D I G L L A A M L S G N M M H D A Q L V T L D A F					
		220	230	240	250	260	270

40 The complete length ORF105a nucleotide sequence <SEQ ID 413> is:

	1	ATGCCGACCG	TCCGTTTTAC	CGAATCCGTC	AGCAAACACG	ACCTTGATGC
	51	CCTATTGAG	TGGGCAAAGG	CAAGTTACGG	TGCGGAAAGT	TGCTGGAAAA
	101	CGCTGTATCT	GAAACGGTCTG	CCTTTGGGCA	ATCTGTGCGC	GGAAATGGGCG
45	151	GAGCGCGTCA	AAAAAGACTG	GGAGGACGGC	TGCTCGGAGT	CTTCAGACGG
	201	CATTTTCCTG	AATGCGGACG	GCTGGCCAGA	TATGGGCAGA	CGCTTGCAGC
	251	ACCTCGCCCG	AATATGAAAA	GAAAGCGGGAC	TGCTTTCACGG	CTGGCGCGAC
	301	GAGTGTTCG	ACCTGACCGA	CGCGGCGCAG	AATCCCTGTG	TCGCGCTCGA
	351	ACGCCCGCCT	TTCGTCGCT	TCGGACTGCT	CAGCCGCGCC	GTTCCATCTCA
50	401	ACGGTTTGGT	CGAATCGGAC	GGCCGATGGC	ATTTCTGGAT	AGGCAGGCGC
	451	AGTCCGCACA	AAGCAGTCGA	TCCCGACAAA	CTCGACAATA	CTGCCCGCCG
	501	CGGTGTTTCC	AGCGGTGAAT	TGCCGTCTGA	AACCGTGTGT	CGCGAAAGCA
	551	GCGAAGAAGC	CGGTTTGGAT	AAAACGCTGC	TTCCGCTCAT	CCGCCCGGTA
	601	TCGCAGCTGC	ACAGCCTGCG	CCCCGTGAGC	CGGGGTGTGC	ACAATGAAAT
55	651	CCTGTATGTA	TTCGATGCGC	TCCTGCCCGA	AACCTTCTGT	CCTGAAATC
	701	AGGATGGCGA	AGTGCGGGT	TTTGAGAAAA	TGGACATCGG	CGGTCTGTTG
	751	GCTGCCATGT	TGTCGGGAAA	CATGATGCAC	GACGCGCAAC	TGGTTACGCT
	801	GGACGCGTTT	TGCCGTTACG	GTCTGATTGA	TGCCGCCCAT	CCGCTGTCCG
	851	AGTGGCTGGA	CGCGATACTG	TTATAG		

This encodes a protein having amino acid sequence <SEQ ID 414>:

1 MPTVRFTESV SKHDLDALE WAKASYGAES CWKTLYLNGL PLGNLSPEWA  
 51 ERVKKDWEAG CSESSDGIFL NADGWPDMDGR RLQHLARIWK EAGLLHGWRD  
 101 ECFDLTDGGS NPLFALERAA FRPFGLLSRA VHLNGLVESD GRWHFWIGRR  
 151 SPHKAVDPDK LDNTAAGGVS SGELPSETVC RESSEEAGLD KTLPLIRPV  
 201 SQLHSLRPVS RGVHNEILYV FDAVLPETFL PENQDGEVAG FEKMDIGGLL  
 251 AAMLSGNMMH DAQLVTLDAL CRYGLIDAAH PLSEWLDGIR L\*

ORF105a and ORF105-1 show 93.8% identity in 291 aa overlap:

		10	20	30	40	50	60
10	orf105a.pep	MPTVRFTESVSKHDLDALEWAKASYGAESCWKTLYLNGLPLGNLSPEWAERVKKDWEAG					
	orf105-1	MPTVRFTESVSKQDLDALEWAKASYGAESCWKTLYLNGLPLGNLSPEWVERVKDWEAG					
		10	20	30	40	50	60
15	orf105a.pep	70	80	90	100	110	120
		CSESSDGIFLNADGWPDMDGRRLQHLARIWKEAGLLHGWRDECFDLTDGGSNPLFALERAA					
	orf105-1	CSESSDGIFLNADGWPDMDGGRLLQHLALGWHCAGLLDGRNECFDLTDGGGNPLFTLERAA					
		70	80	90	100	110	120
20	orf105a.pep	130	140	150	160	170	180
		FRPFGLLSRAVHLNGLVESDGRWHFWIGRRSPHKAVDPDKLDNTAAGGVSSGELPSETVC					
	orf105-1	FRPFGLLSRAVHLNGLTESDGRWHFWIGRRSPHKAVDPDKLDNTAAGGVSSGEMPSEAVC					
		130	140	150	160	170	180
25	orf105a.pep	190	200	210	220	230	240
		RESSEEAGLDKTLPLIRPVSQLHSLRPVSRGVHNEILYVDAVLPETFLPENQDGEVAG					
	orf105-1	RESSEEAGLDKTLPLIRPVSQLHSLRSVSRGVHNEILYVDAVLPETFLPENQDGEVAG					
		190	200	210	220	230	240
30	orf105a.pep	250	260	270	280	290	
		FEKMDIGGLLAAMLSGNMMHDAQLVTLDALCRYGLIDAAHPLSEWLDGIRLX					
	orf105-1	FEKMDIGGLLDAMLSGNMMHDAQLVTLDALCRYGLIDAAHPLSEWLDGIRLX					
		250	260	270	280	290	

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF105 shows 87.5% identity over a 312aa overlap with a predicted ORF (ORF105.ng) from *N.*

40 *gonorrhoeae*:

	orf105.pep	MVARRAHNPKVVGSNPXATXFTQPRFNAEXVLXLPVSCFLFPAASVFCRIFLPAAISER	60
	orf105ng	MVARRAHNPKVVGSNPAPATKYQTPRFNAEGLVLF-----FLFPAASVFCRIFLPAAISER	55
45	orf105.pep	QTAVCLRLQIQAVWLQSSALSSRKPTMPTVRFTESVSKQDLDALEWAKASYGAESCWKT	120
	orf105ng	QAAVCLRLQIQAVWLQSSALCSRKPAMPTVRFTESVSKQDLDALEWAKASYGAESCWKT	115
50	orf105.pep	LYLNGXPLGNLSPEWVERVKKDWEAGCSESSDGIFLNADGWPDMDGGRLLQHLALGWHCAGL	180
	orf105ng	LYLNRLPLGNLSPEWAERIKDWEAGCSESSNGIFLNADGWPDMDGGRLLQHLARTWNKAGL	175
	orf105.pep	LDGWRNECFDLTDGGGNPLFTLERAXRXPGLLSRAVHLNGLTESDGRWHFWIGRRSPHK	240
55	orf105ng	LHGWRNECFDLTDGGGNPLFTLERAAFRPFGLLIRAVHLNGLVESNGRWHFWIGRRSPHK	235
	orf105.pep	AVDPNKLNTXAGGVSGGEMPSEAVCRESEEAGLDKTLPLIRPVSQLHSLRSVSRGVH	300
	orf105ng	AVDPGKLDNIAGGGVSGGEMPSEAVCRESEEAGLDKTLPLIRPVSQLHSLRSVSRGVH	295
60	orf105.pep	NEILYVDAVLP	312
	orf105ng	NEILYVDAVLPETFLPENQDGEVAGFEKMDIGGLLDAMLSKNMMHDAQLVTLDALFYRYG	355

A complete length ORF105ng nucleotide sequence <SEQ ID 415> was predicted to encode a protein having amino acid sequence <SEQ ID 416>:

```

1  MVARRAHNPK VVGSNPAPAT KYOTPRFNAE GVLFFLFPA SVFCRIFLPA
5  51  AISERQAAVC LRLQIQAVWL QSSALCSRKP AMPTVRFETES VSKQDLDALE
101 ERAKASYGAE SCWKTLYLNR LPLGNLSPEW AERIKKDWEA GCSESSNGIF
151 LNADGWPDMM GRLQHLARTW NKAGLLHGWR NECFDLTDGG GNPLFTLERA
201 AFRPFGLLIR AVHLNGLVES NGRWHFWIGR RSPHKAVDPG KLDNIAGGGV
251 SGGEMPSEAV CRESSEEAGL DKTFLPLIRP VSRHLSLRPV SRGVHNEILY
301 VFDAVLPETF LPENQDGEVA GFEMDIGGL LDAMLSKNMM HDAQLVTLDA
10 351 FYRYGLIDAA HPLSEWLDGI RL*

```

Further work revealed the complete nucleotide sequence <SEQ ID 417>:

```

1  ATGCCGACCG TCCGTTTTAC CGAATCCGTC AGCAAACAAG ACCTTGATGC
5  51  CCTGTTCGAG CGGGCAAAAG CAAGTTACGG TGCCGAAAGT TGCTGGAAAA
15  101 CGCTGTATCT GAACCGTCTT CCTTTGGGCA ATCTGTGCGC GGAATGGGCT
151 GAGCGCATCA AAAAAGACTG GGAGGCAGGC TGCTCCGAGT CTTAGACGCG
201 CATTTTCTG AATGCGGACG GCTGGCCGGA TATGGGCGGA CGCTTGACAGC
251 ACCTCGCCCG CACATGGAAC AAGGCGGGGC TGCTTCACGG ATGGCGCAAC
301 GAGTGTTTCG ACCTGACCGA CGGCGGCGGC AACCCCTTGT TCACGCTCGA
20 351 ACGCGCCGCT TTCCGTCGCT TCGGACTACT CAGCCGCGCC GTCCATCTCA
401 ACGGTTTGGT CGAATCGAAC GGCAGATGGC ATTTTGGAT AGGCAGGCGC
451 AGTCCGCACA AAGCAGTCGa tcCCGGCAAG CTGACAATA TTGCCGCGCG
501 CGGTGTTTCC GGCGGCGAAA TGCCGTCTGA AGCCGTGTGC CGCGAAAGCA
551 GCGAAGAAGC CGGTTTGGAT AAAACGCTGT TTCCGCTCAT CCGCCCAGTA
601 TCGCGGCTGC ACAGCCTTCG CCCCGTCAGC CGAGGTGTGC ACAATGAAAT
25 651 CCTGTATGTG TTCGATGCCG TCCTGCCCGA AACCTTCCTG CCTGAAATC
701 AGGATGGCGA GGTAGCGGGT TTTGAAAAGA TGGACATTGG CGGCCTATTG
751 GATGCCATGT TGTCGAAAAA CATGATGCAC GACGCGCAAC TGGTTACGCT
801 GGACGCGTTT TACCGTTACG GTCTGATTGA TGCCGCCCAT CCGCTGTCCG
851 AGTGGCTGGA CGGCATACGT TTATAG

```

This corresponds to the amino acid sequence <SEQ ID 418; ORF105ng-1>:

```

1  MPTVRFETESV SKQDLDALE RAKASYGAES CWKTLYLNRL PLGNLSPEWA
5  51  ERIKKDWEAG CSESSDGIFL NADGWPDMMG RLQHLARTWN KAGLLHGWRN
101 ECFDLTDGGG NPLFTLERA FRPFGLLSRA VHLNGLVESN GRWHFWIGRR
151 SPHKAVDPGK LDNIAGGGVS GGEMPSEAVC RESSEEAGLD KTLFPLIRPV
201 SRLHSLRPVS RGVHNEILYV FDAVLPETFL PENQDGEVAG FEKMDIGGLL
35 251 DAMLSKNMMH DAQLVTLDAF YRYGLIDAAH PLSEWLDGIR L*

```

ORG105ng-1 and ORF105-1 show 93.5% identity in 291 aa overlap:

```

10      20      30      40      50      60
40  orf105-1.pep  MPTVRFETESVSKQDLDALEWAKASYGAESCWKTLYLNGLPLGNLSPEWVERVKDWEAG
    orf105ng-1  MPTVRFETESVSKQDLDALEWAKASYGAESCWKTLYLNRLPLGNLSPEWAERIKDWEAG
10      20      30      40      50      60

70      80      90      100     110     120
45  orf105-1.pep  CSESSDGIFLNADGWPDMMGRLQHLALGWHCAGLLDGWRNECFDLTDGGGNPLFTLERA
    orf105ng-1  CSESSDGIFLNADGWPDMMGRLQHLARTWNKAGLLHGWRNECFDLTDGGGNPLFTLERA
70      80      90      100     110     120

130     140     150     160     170     180
50  orf105-1.pep  FRPFGLLSRAVHLNGLTESDGRWHFWIGRRSPHKAVDPNKLNTAAGGVSGGEMPSEAVC
    orf105ng-1  FRPFGLLSRAVHLNGLVESNGRWHFWIGRRSPHKAVDPGKLDNIAGGGVSGGEMPSEAVC
130     140     150     160     170     180

190     200     210     220     230     240
55  orf105-1.pep  RESSEEAGLDKTLPLIRPVSQHLRLSVSRGVHNEILYVFDAVLPETFLPENQDGEVAG
    orf105ng-1  RESSEEAGLDKTLFPLIRPVSRLHSLRPVSRGVHNEILYVFDAVLPETFLPENQDGEVAG
190     200     210     220     230     240
60

```

```

                250      260      270      280      290
orf105-1.pep  FEKMDIGGLLDAMLSGNMMHDAQLVTLDAFCRYGLIDAAHPLSEWLDGIRLX
               |||||
orf105ng-1    FEKMDIGGLLDAMLSKNMMHDAQLVTLDAFYRYGLIDAAHPLSEWLDGIRLX
                250      260      270      280      290

```

Furthermore, ORF105ng-1 shows homology with a yeast enzyme:

```

sp|P41888|TNR3_SCHPO THIAMIN PYROPHOSPHOKINASE (TPK) (THIAMIN KINASE)
>gi|1076928|pir||S52350 thiamin pyrophosphokinase (EC 2.7.6.2) - fission yeast
(Schizosaccharomyces pombe) >gi|666111 (X84417) thiamin pyrophosphokinase
[Schizosaccharomyces pombe] >gi|2330852|gnl|PID|e334056 (Z98533) thiamin
pyrophosphokinase [Schizosaccharomyces pombe] Length = 569
Score = 105 bits (259), Expect = 4e-22
Identities = 64/192 (33%), Positives = 94/192 (48%), Gaps = 3/192 (1%)

```

```

Query: 268 NKAGLLHGWRNECFDLTDGGGNPLFTLERAAPFRPGLLSRAVHLNGLVESNGRW--HFWI 441
      N  G+  WRNE + +      P+ +ER  F  FG LS  VH  + +      W+
Sbjct: 96  NTFGIADQWRNELYTVYGSKKPVLAVERGGFWLFGFLSTGVHCTMYIPATKEHPLRIWV 155

```

```

Query: 442 GRRSPHKAVDPGKLDNIAGGGVSGGEMPSEAVCRESEEAGLDKTLFPLIRPVSRLHSLR 621
      RRSP K  P  LDN  GG++ G+  + +E SEEA LD +  LI P  + ++
Sbjct: 156 PRRSPTKQTWPNYLDNSVAGGIAHGDSVIGTMIKEFSEANLDVSSMNI-PCGTVSYIK 214

```

```

Query: 622 PVSRG-VHNEILYVFDAVLPETFLPENQDGEVAGFEKMDIGGLLDAMLSKNMMHDAQLVT 798
      R  +  E+ YVFD + + +P  DGEVAGF + + +L  +  K+  +  LV
Sbjct: 215 MEKRHWIQPELQYVFDLPVDDLVIPRINDGEVAGFSLPLNQVLHELELKSFKPNCALVL 274

```

```

Query: 799 LDAFYRYGLIDAAHP 843
      LD  R+G+I  HP
Sbjct: 275 LDFLIRHGIITPQHP 289

```

Based on this analysis, including the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 49

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 419>:

```

1  ATGAATAGAC CCAAGCAACC CTTCTTCCGT CCCGAAGTCG CCGTTGCCCG
51  CCAAACCAGC CTGACGGGTA AAGTGATTCT GACACGACCG TTGTCATTTT
101 CCCTATGGAC GACATTTGCA TCGATATCTG CGTTATTGAT TATCCTGTTT
151 TTGATATTTG GTAACATATC GCGAAAGACA ACAGTGGAGG GACAAATTTT
201 ACCTGCATCG GCGTAATCA GGGTGTATGC ACCGgATACG rGkACAATTA
251 CAGCGAAATT CGTGGAAGAT GGmsAAAAGG TTAAGGCTGG CGACAAGCTA
301 TTTGCGCTTT CGACCTCACG TTTGCGCGCA GGAGGTAGCG TGCAGCAGCA
351 GTTGAAAACG GAGGCAGTTT TGAAGAAAAC GTTGGCAGAA CAGGAAGCTG
401 GTCGTCTGAA GCTGATACAC GGGAAATGAAA CGCGCAgCcT TAAAGCAACT
451 GTCGAACGTT TGGAAAACCA GGAACCTCCAT ATTTCGCAAC AGATAGACGG
501 TCAGAAAAGG CGCATTAGAC TTGCGGAAGA AATGTTGCAG AAATATCGTT
551 TCCTATCCGC .CAATGA

```

This corresponds to the amino acid sequence <SEQ ID 420; ORF107>:

```

1  MNRPKQFFFR PEVAVARQTS LTGKVILTRP LSFSLWTTFA SISALLIILF
51  LIFGNYTRKT TVEGQILPAS GVIRVYAPDT XTITAKEVED GXKVKAGDKL
101 FALSTSRFGA GGSVQQQLKT EAVLKKTLE QELGRLKLIH GNETRSLKAT
151 VERLENQELH ISQQIDGQKR RIRLAEEMLQ KYRFLSXQ*

```

Computer analysis of this amino acid sequence gave the following results:



Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF107 shows 97.8% identity over a 186aa overlap with an ORF (ORF107a) from strain A of *N.*

*meningitidis*:

5	orf107.pep	10 20 30 40 50 60	MNRPKQFFRPEVAVARQTSLTGKVLTRPLSFSLWTTFASISALLIILFLIFGN YTRKT
	orf107a	10 20 30 40 50 60	MNRPKQFFRPEVAVARQTSLTGKVLTRPLSFSLWTTFASISALLIILFLIFGN YTRKT
10	orf107.pep	70 80 90 100 110 120	TVEGQILPASGVIRVYAPDXTITAKFVEDGXKVKAGDKLFALSTSRFGAGGSVQQQLKT
	orf107a	70 80 90 100 110 120	TVEGQILPASGVIRVYAPDGTITAKFXEDGEKVKAGDKLFALSTSRFGAGDSVQQQLKT
15	orf107.pep	130 140 150 160 170 180	EAVLKKTAEQELGRLKLIHGNETRSLKATVERLENQELHISQQIDGQKRIRLAEEMLO
	orf107a	130 140 150 160 170 180	EAVLKKTAEQELGRLKLIHGNETRSLKATVERLENQELHISQQIDGQKRIRLAEEMLO
20	orf107.pep	189	KYRFLSXQX
	orf107a	190 200 210 220 230	KYRFLSANDAVPKQEMMNKAEELLEQKAKLDAYRREEVGLLQEIRTQNLTLXSLPQAAX

The complete length ORF107a nucleotide sequence <SEQ ID 421> is:

30	1	ATGAATAGAC CCAAGCAACC NTTCTTCCGT CCCGAAGTCG CCGTTGCCCG
	51	CCAAACCAGC CTGACGGGTA AAGTGATTCT GACACGACCG TTGTCATTTT
35	101	CCCTATGGAC GACATTTGCA TCGATATCTG CGTTATTGAT TATCCTGTTT
	151	TTGATATTG GTAACATATAC GCGAAAGACA ACAGTGGAGG GACAAATTTT
40	201	ACCTGCATCG GCGGTAATCA GGGTGTATGC ACCGGATACG GGGACAATTA
	251	CNGCGAAATT CNTGGAAGAT GGAGAAAAGG TTAAGGCTGG CGACAAGCTA
45	301	TTTGCCTTT CGACCTCACG TTTCCGCGCA GGAGATAGCG TGCAGCAGCA
	351	GTTGAAAACG GAGGCAGTTT TGAAGAAAAC GTTGGCAGAA CAGGAACCTG
50	401	GTCGTCTGAA GCTGATACAC GGGAAATGAAA CGCGCAGCCT TAAAGCAACT
	451	GTCGAACGTT TGGAAAACCA GGAATCCAT ATTTCGCAAC AGATAGACCG
55	501	TCAGAAAAGG CGCATTAGAC TTGCGGAAGA AATGTTGCAG AAATATCGTT
	551	TCCTATCCGC CAATGATGCA GTGCCAAAAC AAGAAATGAT GAATGTCAAG
60	601	GCAGAGCTTT TAGAGCAGAA AGCCAACTT GATGCCTACC GCCGAGAAGA
	651	AGTCGGGCTG CTTCAAGAAA TCCGCACGCA GAATCTGACA TTGGNNAGCC
65	701	TCCCCAAGC GGCATGA

This encodes a protein having amino acid sequence <SEQ ID 422>:

45	1	MNRPKQFFR PEVAVARQTS LTGKVLTRP LSFSLWTTFA SISALLIILF
	51	LIFGN YTRKT TVEGQILPAS GVIRVYAPDT GTITAKFXED GEKVKAGDKL
50	101	FALSTSRFGA GDSVQQQLKT EAVLKKTAE QELGRLKLIH GNETRSLKAT
	151	VERLENQELH ISQQIDGQKR RIRLAEEMLO KYRFLSANDA VPKQEMMNVK
55	201	AELLEQKAKL DAYRREEVGL LQEIRTQNL TLXSLPQA*

50 Homology with a predicted ORF from *N.gonorrhoeae*

ORF107 shows 95.7% identity over a 188aa overlap with a predicted ORF (ORF107.ng) from *N.*

*gonorrhoeae*:

55	orf107.pep	MNRPKQFFRPEVAVARQTSLTGKVLTRPLSFSLWTTFASISALLIILFLIFGN YTRKT	60
	orf107ng	MNRPKQFFRPEVAIARQTSLTGKVLTRPLSFSLWTTFASISALLIILFLIFGN YTRKT	60
60	orf107.pep	TVEGQILPASGVIRVYAPDXTITAKFVEDGXKVKAGDKLFALSTSRFGAGGSVQQQLKT	120
		:	

orf107ng TMEGQILPASGVIRVYAPDTGTITAKFVEDGEKVKAGDKLFALSTSRFGAGGSVQQQLKT 120  
 orf107.pep EAVLKKTAEQELGRLKLIHNETRSLKATVERLENQELHISQQIDGQKRRIRLAEEMLQ 180  
 5 orf107ng EAVLKKTAEQELGRLKLIHENETRSLKATVERLENQKLHISQQIDGQKRRIRLAEEMLR 180  
 orf107.pep KYRFLSXQ 188  
 orf107ng KYRFLSAQ 188

10 The complete length ORF107ng nucleotide sequence <SEQ ID 423> is predicted to encode a protein having amino acid sequence <SEQ ID 424>:

1 MNRPKQPFER PEVAIARQTS LTGKVILTRP LSFSLWTTFA SISALLIILF  
 51 LIEGNYTRKT TMEGQILPAS GVIRVYAPDT GTITAKFVED GEKVKAGDKL  
 101 FALSTSRFGA GGSVQQQLKT EAVLKKTAE QELGRLKLIH ENETRSLKAT  
 15 151 VERLENQKLH ISQQIDGQKR RIRLAEEMLR KYRFLSAQ\*

Based on the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 50

20 The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 425>:

1 ATGCTGAATA CTTTTTTTGC CGTATTGGGC GGCTGCCTGC TGCT.TTGCC  
 51 GTGCGGCAAA TCCGTAAATA CGGCGGTACA GCCGCAAAAC GCGGTACAAA  
 101 GCGCGCCGAA ACCGGTTTTT AAAGTCATAT ATATCGACAA TACGGCGATT  
 15 151 GCCGGTTTGG ATTTGGGACA AAGCAGCGAA GGCAAAACCA ACGACGGCAA  
 201 AAAACAAATC AGTTATCCGA TTAAAGGCTT GCCGGAACAA AATGTTATCC  
 251 GACTGATCGG CAAGCATCCC GCGCACTTGG AAGCCGTCAG CGGCAATGT  
 301 ATGGAACCG ATGATAAGGA CAGTCCGCA GGTGGGCAG AAAACGGCGT  
 351 GTGCCATACC TTGTTTGCCA AACTGGTGGG CAATATCGCC GAAGACGGCG  
 10 401 GCAAACTGAC GGATTACCTA GTTTCGCATG CCGCCCTGCA ACCCTATCAG  
 451 GCAGGCAAAA GCGGCTATGC CGCCGTGCAG AACGGACGCT ATGTGCTGGA  
 501 AATCGACAGC GAAGGGGCGT TTTATTTCCT CCGCCGCCAT TATTGA

This corresponds to the amino acid sequence <SEQ ID 426; ORF108>:

1 MLNTFFAVLG GCLLXLPCGK SVNTAVQPQN AVQSAPKPVF KVIYIDNTAI  
 35 51 AGLDLGQSSE GKTNDGKKQI SYPIKGLPEQ NVIRLIGKHP GDLEAVSGKC  
 101 METDDKDSPA GWAENGVCHT LFAKLVGNIA EDGKLTIDYL VSHAALQPYQ  
 151 AGKSGYAAVQ NGRVLEIDS EGAFYFRRRH Y\*

Further work revealed the following DNA sequence <SEQ ID 427>:

1 ATGCTGAAAA CATCTTTTGC CGTATTGGGC GGCTGCCTGC TGCTTGCCGC  
 40 51 CTGCGGCAAA TCCGAAAATA CGGCGGAACA GCCGCAAAAC GCGGTACAAA  
 101 GCGCGCCGAA ACCGGTTTTT AAAGTCAAAT ATATCGACAA TACGGCGATT  
 151 GCCGGTTTGG ATTTGGGACA AAGCAGCGAA GGCAAAACCA ACGACGGCAA  
 201 AAAACAAATC AGTTATCCGA TTAAAGGCTT GCCGGAACAA AATGTTATCC  
 251 GACTGATCGG CAAGCATCCC GCGCACTTGG AAGCCGTCAG CGGCAATGT  
 45 301 ATGGAACCG ATGATAAGGA CAGTCCGCA GGTGGGCAG AAAACGGCGT  
 351 GTGCCATACC TTGTTTGCCA AACTGGTGGG CAATATCGCC GAAGACGGCG  
 401 GCAAACTGAC GGATTACCTA GTTTCGCATG CCGCCCTGCA ACCCTATCAG  
 451 GCAGGCAAAA GCGGCTATGC CGCCGTGCAG AACGGACGCT ATGTGCTGGA  
 501 AATCGACAGC GAAGGGGCGT TTTATTTCCT CCGCCGCCAT TATTGA

50 This corresponds to the amino acid sequence <SEQ ID 428; ORF108-1>:

```

1  MLKTSFAVLG GCLLLAACGK SENTAEQPQN AVQSAPKPVF KVKYIDNTAI
51  AGLDLGQSSE GKTNDGKKQI SYPIKGLPEQ NVIRLIGKHP GDLEAVSGKC
101 METDDKDDPA GWAENGVCHT LFAKLVGNI AEDGKLT DYL VSHAALQPYQ
151 AGKSGYAAVQ NGRYVLEIDS EGAFYFRRRH Y*

```

- 5 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.gonorrhoeae*

ORF108 shows 88.4% identity over a 181aa overlap with a predicted ORF (ORF108.ng) from *N.gonorrhoeae*:

```

10  orf108.pep  MLNTFFAVLGGCLLLXPCGKSVNTAVQPQNAVQSAPKPVFKVIYIDNTAIAGLDLGQSSE  60
      ||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      orf108ng  MLKIPFAVLGGCLLLAACGKSENTAEQPQNAASAPKPVFKVKYIDNTAIAGLALGQSSE  60

      orf108.pep  GKTNDGKKQISYPIKGLPEQNVIRLIGKHPGDLEAVSGKCMETDDKDSPAGWAENGVCHT  120
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
15  orf108ng  GKTNDGKKQISYPIKGLPEQNAVRLTGKHPNDLEAVVGKCMETDGKDAPSGWAENGVCHT  120

      orf108.pep  LFAKLVGNI AEDGKLT DYL VSHAALQPYQAGKSGYAAVQNGRYVLEIDSEGA FYFRRRH Y  181
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
20  orf108ng  LFAKLVGNI AEDGKLT DYL ISHSALQPYQAGKSGYAAVQNGRYVLEIDSEGA FYFRRRH Y  181

```

ORF108-1 shows 92.3% identity with ORF108ng over the same 181 aa overlap:

```

25  orf108-1.pep  MLKTSFAVLGGCLLLAACGKSENTAEQPQNAVQSAPKPVFKVKYIDNTAIAGLDLGQSSE  60
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      orf108ng-1  MLKIPFAVLGGCLLLAACGKSENTAEQPQNAASAPKPVFKVKYIDNTAIAGLALGQSSE  60

      orf108-1.pep  GKTNDGKKQISYPIKGLPEQNVIRLIGKHPGDLEAVSGKCMETDDKDSPAGWAENGVCHT  120
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      orf108ng-1  GKTNDGKKQISYPIKGLPEQNAVRLTGKHPNDLEAVVGKCMETDGKDAPSGWAENGVCHT  120

30  orf108-1.pep  LFAKLVGNI AEDGKLT DYL VSHAALQPYQAGKSGYAAVQNGRYVLEIDSEGA FYFRRRH Y  181
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      orf108ng-1  LFAKLVGNI AEDGKLT DYL ISHSALQPYQAGKSGYAAVQNGRYVLEIDSEGA FYFRRRH Y  181

```

The complete length ORF108ng nucleotide sequence <SEQ ID 429> is:

```

35  1  ATGCTGAAAA tacctTTTGC CGTGTtgggc ggCtgcctGC TGCTTGCCGC
      51  CTGCGGCAAA TCCGAAAATA cggcggaACA GCCGCAAAAT gcggCACAAA
      101  GCGCGCCGAA ACCGGTTTTC AAAGTCAAAT ACATCGACAA TACGGCGATT
      151  GCCGGTTTGG CTTTGGGACA AAGTAGCGAA GGCAAAACCA acgacgGCAA
      201  AAAACAAATC AGTTATccgA TTAAAGGCTT GCCGGAACAA Aacgccgtcc
40  251  gGCTGACCGG AAAGCATCCC AACGACTTGG AagccgtcgT CGGCAAATGT
      301  ATGGAAACCG ACGGAAAGGA CGCGCCTTCG GGCTGGGCGG AAAACGGCGT
      351  GTGCCATACC TTGTTTGCCA AACTGGTGGG CAATATCGCC GAAGACGGCG
      401  GCAAACTGAC TGATTACCTG ATTTGCGATT CCGCCCTGCA ACCCTATCAG
      451  GCAGGCAAAA GCGGCTATGC CGCCGTGCAG AACGGACGCT ATGTGCTGGA
      501  AATCGACAGC GaggggGCGT TTTATttccg ccgccgceat tattgA

```

- 45 This encodes a protein having amino acid sequence <SEQ ID 430>:

```

1  MLKIPFAVLG GCLLLAACGK SENTAEQPQN AAQSAPKPVF KVKYIDNTAI
51  AGLALGQSSE GKTNDGKKQI SYPIKGLPEQ NAVRLTGKHP NDLEAVVGKC
101 METDGKDAPS GWAENGVCHT LFAKLVGNI AEDGKLT DYL ISHSALQPYQ
151 AGKSGYAAVQ NGRYVLEIDS EGAFYFRRRH Y*

```

- 50 Based on this analysis, including the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site (underlined) and a putative ATP/GTP-binding site motif A (P-loop, double-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and

*N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 51

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 431>:

```

5      1  ATGGAAGATT TATATATAAT ACTCGCTTTG GGTTCGGTTG CGATGATTGC
      51  CGGATTTATC GATgcatTg cGgCGGGGG TGGTTTGATT ACGCTGCCCCG
     101  CACTCTTGTT GGCAGGTATT CCTCCCGTGT CGGCAATTGC CACCAACAAG
     151  CTGCAAGCAG CCGCTGCTAC GTTTTCAGCT ACGGTTTCTT TTGCACGCAA
     201  AGGTTTGATT GATTGGAAGA AAGGTCTCCC GATTGCCGCA GCATCGTTTG
     251  TAGGCGGCGT GGcCGGTGCA TTATCGGTCA GCTTGGTTTC CAAAGATATT
     301  CTgCTgGCGG TCGTGCCGGT TTTGTTGATA TTTGTGCGAC TGTATTTTGT
     351  GTTTTCGCCC AAGCTCGACG GCAGTAAGGA AGGCAAAGCC AGAATGTCTT
     401  TTTTCTGTGT cGGGCTGACG GTCGC.ACCG CTTTGGGTT TTTACGACGG
     451  TGTGTTTCGA CCGGGTGTGCG GCTCGTTTTT TCTGATTGCC TTTATTGTTT
    15  501  TGCTCGGCTG CAAGCTGTTG AACGCGATGT CTTACACCAA ATGGCGAAC
     551  GTTGCCGTGCA ATCTTGGTTC GCTATCGGTA TTCCTGCTGC ACGGTTTCGAT
     601  TATTTTCCCG ATTCGCGCAA CGaTGGCGGT CGGTGCGTTT GTCGGTGCGA
     651  ATTTAgGTGC GAGATTTGCC GTaCgctTCG GTTCGAAGCT GATTAA
  
```

This corresponds to the amino acid sequence <SEQ ID 432; ORF109>:

```

20      1  MEDLYIILAL GLVAMIAGFI DAIAGGGGLI TLPALLLAGI PPVSAIATNK
      51  LQAAAATFSA TVSFARKGLI DWKKGLPIAA ASFVGGVAGA LSVSLVSKDI
     101  LLAVVPVLLI FVALYFVFSP KLDGSKEGKA RMSFFLFGLT VXTAFGFLRR
     151  CVRTGCRLVF SDCLYCFARL QAVERDVLHQ IGERCLQSWF AIGIPAARFD
     201  YFPDCGNDGG RCVCRCEFRG EICRTLRFEA D*
  
```

Further work revealed the following DNA sequence <SEQ ID 433>:

```

30      1  ATGGAAGATT TATATATAAT ACTCGCTTTG GGTTCGGTTG CGATGATTGC
      51  CGGATTTATC GATGCGATTG CCGGCGGGGG TGGTTTGATT ACGCTGCCCCG
     101  CACTCTTGTT GGCAGGTATT CCTCCCGTGT CGGCAATTGC CACCAACAAG
     151  CTGCAAGCAG CCGCTGCTAC GTTTTCAGCT ACGGTTTCTT TTGCACGCAA
     201  AGGTTTGATT GATTGGAAGA AAGGTCTCCC GATTGCCGCA GCATCGTTTG
     251  TAGGCGGCGT GGCCGGTGCA TTATCGGTCA GCTTGGTTTC CAAAGATATT
     301  CTGCTGGCGG TCGTGCCGGT TTTGTTGATA TTTGTGCGAC TGTATTTTGT
     351  GTTTTCGCCC AAGCTCGACG GCAGTAAGGA AGGCAAAGCC AGAATGTCTT
     401  TTTTCTGTGT CCGGCTGACG GTCGCACCGC TTTTGGGTTT TTACGACGGT
     451  GTGTTTCGGAC CCGGTGTGCG CTCGTTTTTT CTGATTGCCT TTTATTGTTT
     501  GCTCGGCTGC AAGCTGTTGA ACGCGATGTC TTACACCAA TTGGCGAACG
     551  TTGCCTGCAA TCTTGGTTTC CTATCGGTAT TCCTGCTGCA CGGTTTCGATT
     601  ATTTTCCCGA TTGCGGCAAC GATGGCGGTC GGTGCGTTTG TCGGTGCGAA
     651  TTTAGGTGCG AGATTTGCCG TCCGCTTCGG TTCGAAGCTG ATTAAGCCGC
    40  701  TGCTGATTGT CATCAGCATT TCGATGGCTG TGAAATTGTT GATAGACGAG
     751  AGAAATCCGC TGTATCAGAT GATTGTTTCG ATGTTTTAA
  
```

This corresponds to the amino acid sequence <SEQ ID 434; ORF109-1>:

```

45      1  MEDLYIILAL GLVAMIAGFI DAIAGGGGLI TLPALLLAGI PPVSAIATNK
      51  LQAAAATFSA TVSFARKGLI DWKKGLPIAA ASFVGGVAGA LSVSLVSKDI
     101  LLAVVPVLLI FVALYFVFSP KLDGSKEGKA RMSFFLFGLT VAPLLGFYDG
     151  VFGPGVGSFF LIAFIVLLGC KLLNAMSYTK LANVACNLGS LSVFLHGSF
     201  IFPIAATMAV GAFVGANLGA RFAVRFGSKL IKPLLIVISI SMAVKLLIDE
     251  RNPLYQMIVS MF*
  
```

Computer analysis of this amino acid sequence gave the following results:

50 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF109 shows 95.9% identity over a 147aa overlap with an ORF (ORF109a) from strain A of *N.meningitidis*:

		10	20	30	40	50	60
5	orf109.pep	MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA					
	orf109a	MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA					
		10	20	30	40	50	60
10	orf109.pep	TVSFARKGLIDWKKGLPIAAASFVGGVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSP					
	orf109a	TVSFARKGLIDWKKGLPIAAASFAGGVVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSP					
		70	80	90	100	110	120
15	orf109.pep	KLDGSKEGKARMSFFLFGLTVXTAFGLRRCVRTGCRLVFSDCLYCFARLQAVERDVLHQ					
	orf109a	KLDGSKEGKARMSFFLFGLTVAPLLGFYDGVFGPGVGSFFLIAFIVLLGCKLLNAMSYTK					
		130	140	150	160	170	180

The complete length ORF109a nucleotide sequence <SEQ ID 435> is:

20	1	ATGGAAGATT	TATACATAAT	ACTCGCTTTG	GGTTTGGTTG	CGATGATTGC
	51	CGGATTTATC	GATGCGATTG	CGGGTGGGGG	TGGTTTGATT	ACGCTGCCTG
	101	CACTCTTGTT	GGCAGGTATT	CCTCCCGTGT	CGGCAATTGC	CACCAACAAG
	151	CTGCAAGCAG	CCGCTGCTAC	GTTTTCGGCT	ACGGTTTCTT	TTGCACGCAA
	201	AGGTTTGATT	GATTGGAAGA	AAGTCTCCC	GATTGCGGCA	GCATCGTTTG
	251	CAGGCGGCGT	GGTCGGTGCA	TTATCGGTCA	GCTTGGTTTC	CAAAGATATT
25	301	CTGCTGGCGG	TCGTGCCGGT	TTTGTGATA	TTTGTGCGCG	TGTATTTTGT
	351	GTTTTCGCCC	AAGCTCGACG	GCAGTAAGGA	AGGCAAAGCC	AGAATGTCTT
	401	TTTTTCTGTT	CGTCTGACG	GTTGCACCAC	TTTGGGTTT	TTACGACGGT
	451	GTGTTCCGAC	CGGGTGTCCG	CTCGTTTTTT	CTGATTGCCT	TTATTGTTTT
30	501	GCTCGGCTGC	AAGCTGTTGA	ACGCGATGTC	TTACACCAA	TTGGCGAACG
	551	TTGCCTGCAA	TCTTGGTTCG	CTATCGGTAT	TCCTGCTGCA	CGGTTCCGATT
	601	ATTTTCCCGA	TTGCGGCAAC	GATGGCGGTC	GGTGCCTTTG	TCGGTGCGAA
	651	TTTAGGTGCG	AGATTGCCG	TCCGCTTCGG	TTGGAAGCTG	ATTAAGCCGC
	701	TGCTGATTGT	CATCAGCATT	TCGATGGCTG	TGAAATTGTT	GATAGACGAG
	751	AGAAATCCGC	TGTATCAGAT	GATTGTTTCG	ATGTTTTTAA	

35 This encodes a protein having amino acid sequence <SEQ ID 436>:

40	1	MEDLYIILAL	GLVAMIAGFI	DAIAGGGGLI	TLPALLLAGI	PPVSAIATNK
	51	LQAAAATFSA	TVSFARKGLI	DWKKGLPIAA	ASFAGGVVGA	LSVSLVSKDI
	101	LLAVVPVLLI	FVALYFVFSP	KLDGSKEGKA	RMSFFLFGLT	VAPLLGFYDG
	151	VFGPGVGSFF	LIAFIVLLGC	KLLNAMSYTK	LANVACNLGS	LSVFLHGSII
	201	IFPIAATMAV	GAFVGANLGA	RFAVRFGSKL	IKPLLIVISI	SMAVKLLIDE
	251	RNPYQMIVS	MF*			

ORF109a and ORF109-1 show 99.2% identity in 262 aa overlap:

45	orf109a.pep	MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA					
	orf109-1	MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA					
		10	20	30	40	50	60
50	orf109a.pep	TVSFARKGLIDWKKGLPIAAASFAGGVVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSP					
	orf109-1	TVSFARKGLIDWKKGLPIAAASFVGGVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSP					
		70	80	90	100	110	120
55	orf109a.pep	KLDGSKEGKARMSFFLFGLTVAPLLGFYDGVFGPGVGSFFLIAFIVLLGCKLLNAMSYTK					
	orf109-1	KLDGSKEGKARMSFFLFGLTVAPLLGFYDGVFGPGVGSFFLIAFIVLLGCKLLNAMSYTK					
		130	140	150	160	170	180
60	orf109a.pep	LANVACNLGSLSVFLLHGSIIIFPIAATMAVGAFVGANLGARFAVRFGSKLIKPLLIVISI					
		190	200	210	220	230	240

orf109-1	LANVACNLGSLSVFLLHGSII	FPIAATMAVGAFV	GANLGARFAVRFGSKLIKPL	LIVISI
	190	200	210	220 230 240
5	orf109a.pep	250 260		
		SMAVKLLIDERNPLYQMIVSMFX		
	orf109-1	SMAVKLLIDERNPLYQMIVSMFX		
		250 260		

10 Homology with a predicted ORF from *N.gonorrhoeae*

ORF109 shows 98.3% identity over a 231aa overlap with a predicted ORF (ORF109.ng) from *N.*

*gonorrhoeae*:

orf109.pep	MEDLYIILALGLVAMIAGFIDA	IAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA	60
15	orf109ng	MEDLYIILALGLVAMIAGFIDA	IAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA 60
orf109.pep	TVSFARKGLIDWKKGLPIAA	ASFAGGVVGA LSVSLVSKDILLAVVPVLLIFVALYFVFSP	120
20	orf109ng	TVSFARKGLIDWKKGLPIAA	ASFAGGVVGA LSVSLVSKDILLAVVPVLLIFVALYFVFSP 120
orf109.pep	KLDGSKEGKARMSFFLEFLT	TVXTAFGLRRCVRTGCRLVFS	SDCLYCFARLQAVERDVLHQ 180
25	orf109ng	KLDGSKEGKARMSFFLEFLT	TVXTAFGLRRCVRTGCRLVFS
orf109.pep	IGERCLQSWFAIGIPAARFDY	FPDCGNDGGRCVRCCEFRCEICRTLRF	FEAD 231
orf109ng	IGERCLQSWFAIGIPAARFDY	FPDCGNDGGRCVRCCEFRCEICRPLRF	FEAD 231

An ORF109ng nucleotide sequence <SEQ ID 437> was predicted to encode a protein having amino acid sequence <SEQ ID 438>:

30	1	MEDLYIILAL GLVAMIAGFI	DAIAGGGGLI TLPALLLAGI	PPVSAIATNK
	51	LQAAAATFSA TVSFARKGLI	DWKKGLPIAA ASFAGGVVGA	LSVSLVSKDI
	101	LLAVVPVLLI FVALYFVFSP	KLDGSKEGKA RMSFFLEFLT	VATAFGFLRR
	151	CVRTGCRLVF SDCLYCFARL	QAVERDVLHQ IGERCLQSWF	AIGIPAARFD
	201	YFPDCGNDGG RCVRCCEFR	C EICRPLRFEA D*	

## 35 Further work revealed the following gonococcal DNA sequence &lt;SEQ ID 439&gt;:

40	1	ATGGAAGATT TATACATAAT	ACTCGCTTTG GGTGTTGGTTG	CGATGATCGC
	51	CGGATTTATC GATGCGATTG	CGGGCGGGGG TGGTTGATT	ACGCTGCCTG
	101	CACTCTTGTT GGCAGGTATT	CCTCCCGTGT CGGCAATTGC	CACCAACAAG
	151	CTGCAAGCAG CCGCTGCTAC	GTTTTCGGCT ACGGTTTCTT	TTGCACGCAA
	201	AGGTTTGATT GATTGGAAGA	AAGGTCTCCC GATTGCCGCA	GCATCGTTTG
	251	CAGGCGGCGT GGTGCGTGCA	TTATCGGTCA GCTTGGTTTC	CAAAGATATT
	301	TTGCTGGCGG TCGTGCCGGT	TTTGTGATA TTTGTCGCGC	TGTATTTTGT
	351	GTTTTCGCCC AAGCTCGACG	GCAGTAAGGA AGGCAAAGCC	AGAATGTCTT
	401	TTTTTCTATT CGGGCTGACG	GTTGCACCGC TTTTGGGTTT	TTACGACGGT
	451	GTGTTCCGAC CGGGTGTCGG	CTCGTTTTTT CTGATTGCCT	TTATTGTTTT
	501	GCTCGGCTGC AAGCTGTTGA	ACGCGATGTC TTACACCAA	TTGGCGAAGC
	551	TTGCTTGCAA TCTTGGTTCG	CTATCGGTAT TCCTGCTGCA	CGGTTTCGATT
	601	ATTTTCCCGA TTGTGGCAAC	GATGGCGGTC GGTGCGTTTG	TCGGTGCGAA
	651	TTTAGGTGCG AGATTTGCCG	TCCGCTTCGG TTCGAAGCTG	ATTAAGCCGC
	701	TGCTGATTGT CATCAGCATT	TCGATGGCTG TGAAATTGTT	GATAGACGAG
	751	AGAAATCCGC TGTATCAGAT	GATTGTTTCG ATGTTTTAA	

This corresponds to the amino acid sequence <SEQ ID 440; ORF109ng-1>:

55	1	MEDLYIILAL GLVAMIAGFI	DAIAGGGGLI TLPALLLAGI	PPVSAIATNK
	51	LQAAAATFSA TVSFARKGLI	DWKKGLPIAA ASFAGGVVGA	LSVSLVSKDI
	101	LLAVVPVLLI FVALYFVFSP	KLDGSKEGKA RMSFFLEFLT	VAPLLGFYDG
	151	VFGPGVGSFF LIAFIVLLGC	KLLNAMSYTK LANVACNLGS	LSVFLHGSII
	201	IFPIVATMAV GAFVGANLGA	RFAVRFGSKL IKPLLIVISI	SMAVKLLIDE
	251	RNPLYQMIVS MF*		

ORF109ng-1 and ORF109-1 show 98.9% identity in 262 aa overlap:

```

      10      20      30      40      50      60
orfl09ng-1.pep MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAAATFSA
5 orfl09-1      MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAAATFSA
      10      20      30      40      50      60

      70      80      90     100     110     120
orfl09ng-1.pep TVSFARKGLIDWKKGLPIAAASFAGGVVGALSVSLVSKDILLAVVPVLLIFVALYFVFSP
10 orfl09-1      TVSFARKGLIDWKKGLPIAAASFAGGVVGALSVSLVSKDILLAVVPVLLIFVALYFVFSP
      70      80      90     100     110     120

      130     140     150     160     170     180
orfl09ng-1.pep KLDGSKEGKARMSFFLFGTLVAPLLGFYDGVFGPGVGSFFLIAFIVLLGCKLLNAMS YTK
15 orfl09-1      KLDGSKEGKARMSFFLFGTLVAPLLGFYDGVFGPGVGSFFLIAFIVLLGCKLLNAMS YTK
      130     140     150     160     170     180

      190     200     210     220     230     240
orfl09ng-1.pep LANVACNLGSLSVFLLHGSIIFFPIVATMAVGAFVGANLGARFAVRFGSKLIKPLLVISI
20 orfl09-1      LANVACNLGSLSVFLLHGSIIFFPIAATMAVGAFVGANLGARFAVRFGSKLIKPLLVISI
      190     200     210     220     230     240

      250     260
orfl09ng-1.pep SMAVKLLIDERNPLYQMIVSMFX
25 orfl09-1      SMAVKLLIDERNPLYQMIVSMFX
      250     260

```

In addition, ORF109ng-1 shows homology to a hypothetical *Pseudomonas* protein:

```

sp|P29942|YCB9_PSEDE HYPOTHETICAL 27.4 KD PROTEIN IN COBO 3'REGION (ORF9)
>gi|94984|pir||I38164 hypothetical protein 9 - Pseudomonas sp >gi|551929
35 (M62866) ORF9 [Pseudomonas denitrificans] Length = 261
   Score = 175 bits (439), Expect = 3e-43
   Identities = 83/214 (38%), Positives = 131/214 (60%), Gaps = 1/214 (0%)

Query: 41  PPVSAIATNKLQXXXXXXXXXXXXXKRKGLIDWKKGLPIXXXXXXXXXXXXXXXXXXXXKDI 100
      PP+  + TNKLQ                      R+G ++ K+ LP+                      D+
40 Sbjct: 43  PPLQTLGTNKLQGLFGSGSATLSYARRGHVNLKEQLPMALMSAAGAVLGALLATIVPGDV 102

Query: 101 LLAVVPVLLIFVALYFVFSPKLDGSKEGKARMSFFLFGTLVAPLLGFYDGVFGPGVGSFF 160
      L A++P LLI +ALYF  P + G +  +R++ F+F LT+ PL+GFYDGVFGPG GSFF
45 Sbjct: 103 LKAILPFLIIAIALYFGLKPNM-GDVDQHSRVTPEVFTLTLPVPLIGFYDGVFGPGTGSFF 161

Query: 161 LIAFIVLLGCKLLNAMS YTKLANVACNLGSLSVFLLHGSIIFFPIVATMAVGAFVGANLGA 220
      ++ F+ L G  +L A ++TK N  N+G+  VFL  G++++ +  M +G F+GA +G+
50 Sbjct: 162 MLGFVTLAGFGVLKATAHTKFLNFGSNVGAFGVFLFFGAVLWKVGLLMGLGQFLGAQVGS 221

Query: 221 RFAVRFGSKLIKPLLVIVISISMAVKLLIDERNPL 254
      R+A+  G+K+IKPLL+++SI++A++LL D  +PL
55 Sbjct: 222 RYAMAKGAKIIKPLLIVIVSIALAIRLLADPTHPL 255

```

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

**Example 52**

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 441>:

```

1  ..CTGCTAGGGT ATTGCATCGG TTATCGGTAC GGCTGTTGCA GCAAAACCAG
51  CCGCAGACGG ATTATTTGGT CAAATTCGGA TCGTTTTGGG CGAG.ATTTT
101 TGGTTTTCTG GGAAGTATG ACGTCTATGC TTCGGCATGG TTTGTCGTTA
151 TCATGATGTT TTTGGTGGTT TCTACCAGTT TGTGCCTGAT TCGCAATGTG
201 CCGCCGTTCT GCGCGGAAAT GAAGTCTTTT CCGGAAAAGG TTAAGAAAAA
251 ATCTCTGGCG GCGATGCGCC ATTCTTCGCT GTTGGATGTA AAAATTCGCG
301 CCGAGGTTGC CAAACGTTAT CTGGAAGTAC AAGGTTTTCA GGGGAAAACC
351 ATTAACCGTG AAGACGGGTC GGTTCGTATT GCCGCCAAAA AAGGCACAAT
401 GAACAAATGG GGCTATATCT TTGCCCATGT TGCTTTGATT GTCATTTGCC
451 TGGGCGGGTT GATAGACAGT AACCTGCTGT TGAACTGGG TATGCTGACC
501 GGTCGGATTG TTCCGGACAA TCAGGCGGTT TATGCCAAGG ATTTT.AAGC
551 CCGAAAGTAT .TTTGGGTGC gTCCAATCTC TCATTTAGGG GCAACGTCAA
15  601 TATTTCCG.A GGGGCAGAgT GCGGATGTGG TTTTCCTGA

```

This corresponds to the amino acid sequence <SEQ ID 442; ORF110>:

```

1  ..LLGIASVIGT LLQONQPQTD YLVKFGSFWA XIFGFLGLYD VYASAWFVVI
51  MMFLVSTSL CLIRNVPPFW REMKSFREKV KEKSLAAMRH SSLLDVKIAP
101 EVAKRYLEVQ GFQGKTINRE DGSVLIAAKK GTMNKWGYIF AHVALIVICL
151 GGLIDSNLLL KLGMLTGRIF RTIRRFMPRI XKPESXFGCV QSLI*GQRQY
201 FXRGRVMMWF S*

```

Computer analysis of this amino acid sequence gave the following results:

**Homology with ORF88a from *N.meningitidis* (strain A)**

ORF110 shows 91.5% identity over a 188aa overlap with ORF88a from strain A of *N. meningitidis*:

```

25  orf88a.pep  MSKSRSPPLLSPWFAFFSSMRFAVALLSLLGIASVIGTVLQONQPQTDYLVKFGSFWA
    orf110      LLGIASVIGTLLQONQPQTDYLVKFGSFWA
30  orf88a.pep  QIFGFLGLYDVYASAWFVIMMFLVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH
    orf110      XIFGFLGLYDVYASAWFVIMMFLVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH
35  orf88a.pep  SSLLDVKIAPEVAKRYLEVQGFQGKTINREDGSVLIAAKKGTMNKWGYIFAHVALIVICL
    orf110      SSLLDVKIAPEVAKRYLEVQGFQGKTINREDGSVLIAAKKGTMNKWGYIFAHVALIVICL
40  orf88a.pep  GGLIDSNLLLKLGMALTGRIVPDNQAVYAKDFKPESILGASNLSFRGNVNISEGQSADVVF
    orf110      GGLIDSNLLLKLGMALTGRIFRTIRRFMPRIKKPESXFGCVQSLIXGQRQYFXRGRVMMWF
45  orf88a.pep  LNADNGILVQDLPFEVKKLKFHIDFYNTGMPRDFASDIEVTDKATGEKLEKERTIRVNHPLT
    orf110      SX
50

```

However, ORF88 and ORF110 do not align, because they represent two different fragments of the same protein.



Homology with a predicted ORF from *N.gonorrhoeae*

ORF110 shows 88.6% identity over a 211aa overlap with a predicted ORF (ORF110.ng) from *N.gonorrhoeae*:

5	orf110.pep	LLGIASVIGTLLQQNQPTDYLKFGSFWA	30
	orf110ng	MSKSRSISPTLLSRPWFAFFSSMRFAVALLSLLGIASVIGTVLQQNQPTDYLKFGPFWT	60
10	orf110.pep	XIFGFLGLYDVYASAWFVIMMFLVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH	90
	orf110ng	RIFDFLGLYDVYASAWFVIMMFLVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH	120
15	orf110.pep	SSLLDVKIAPEVAKRYLEVQFGQKTINREDGSVLIAAKKGTMNKGWYIFAHVALIVICL	150
	orf110ng	SSLLDVKIAPEVAKRYLEVRFQGTVSREDGSVLIAAKKGTMNKGWYIXAHVALIVICL	180
20	orf110.pep	GGLIDSNLLKLGMLTGRIFRTIRRFMPRIKXPESXFGCVQSLIXGQRQYFXRGRVVMWF	210
	orf110ng	GRLINXNLLKLGMLAGSIFRNNRRVMPRIKXPESIWGGVQSLIKGQRQYFQRGKVRMWF	240
25	orf110.pep	S 211	
	orf110ng	S 241	

The complete length ORF110ng nucleotide sequence <SEQ ID 443> is predicted to encode a protein having amino acid sequence <SEQ ID 444>:

25	1	MSKSRSISPTL	LSRPWFAFFS	SMRFAVALLS	LLGIASVIGT	VLQQNQPTD
	51	YLVKFGPFWT	RIFDFLGLYD	VYASAWFVVI	MMFLVSTSL	CLIRNVPPFW
	101	REMKSFREKV	KEKSLAAMRH	SSLLDVKIAP	EVAKRYLEVR	GFQGTVSRE
	151	DGSVLIAAKK	GTMNKGWYIX	AHVALIVICL	GRLINXNLLL	KLGMLAGSIF
	201	RNNRRVMPRI	SKPESIWGGV	QSLIKGQRQY	FQRGKVRMWF	S*

- 30 Based on the putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

**Example 53**

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 445>:

35	1	ATGCCGTC	CTG	AAACACGC	CT	GCCGAAC	TTT	ATCCGCG	TCT	TGATATT	TGC
	51	CCTGGGT	TTC	ATCTTC	TCTGA	ACGCCT	GTTC	GGAACAA	ACC	GCGCAA	ACCG
	101	TTACCCT	GCA	AGGCGAA	ACG	ATGGGC	CACGA	CCTATAC	CGT	CAAATAC	CCTT
	151	TCAAATA	AATC	GGGACAA	ACT	CCCCTCA	CCT	GCCGAA	ATAC	AAAAAC	GCGAT
40	201	CGATGAC	GCG	CTTAAAG	AAG	TCAACCG	GCA	GATGTCC	ACC	TATCAGC	CCG
	251	ACTCCGA	AAT	CAGCCGG	TTC	AACCAAC	ACA	CAGCCGG	CAG	GCCCTCC	GCG
	301	ATTTCA	AGCG	ACTTCG	CACA	CGTTACT	GCC	GAAGCC	GTC	GCCTGA	ACCG
	351	CCTGAC	ACAC	GGCGCG	CTGG	ACGTAAC	CGT	CGGCCCT	TG	GTCAAC	CTTT
45	401	GGGGATT	TCG	CCCCGAC	AAA	TCCGTTA	CCC	GTGAACC	GTC	GCCGGA	ACAA
	451	ATCAAAC	AGG	CGGCAT	CTTA	TACGGGC	CATA	GACAAAT	CA	TTTGA	AAACA
	501	AGGCAAA	GAT	TACGCT	TCCT	TGAGCAA	AAC	CCACCC	CAAG	GCCTATT	TGG
	551	ATTTAT	CTTC	GATTGCC	AAA	GGCTTC	GCG	TTGATA	AAAGT	TGCGGG	CGAA
50	601	CTGGAAA	AAT	ACGGCAT	TCA	AAATTAT	CTG	GTGAAA	ATCG	GCGGCG	AGTT
	651	GCACGG	CAAA	GGCAAAA	ACG	CGCGCG	GCGA	ACCGTG	GCGC	ATCGGT	ATCG
	701	AGCAGC	CCCA	TATCGT	CCTAA	GGCGG	CAATA	CGCAG	ATTAT	CGTCC	CGCTG
	751	AACAACC	CGTT	CGTTG	CCAC	TTCCG	GCGAT	TACCG	TATTT	TCCAC	GTCGA
	801	TAAAAAC	GCG	AAACGC	CCTCT	CCCATAT	CAT	CAACCC	GAAC	AACAA	ACGAC
	851	CCATCAG	CCA	CAACCT	CGCC	TCCATC	AGCG	TGGT	CGCAG	CAGTG	CGATG
	901	ACGCGG	GACG	GCTTGT	CCAC	AGGATT	ATTC	GTATTG	GCGC	AAACCG	AAAGC
	951	CTTAAAG	CTG	GCAGAG	C	AAAAA	CTCG	TGTTTC	CTG	ATTGTC	AGGG

1001 ATAAAGGCGG CTACCGCACC GCCATGTCTT CCGAATTGA AAAACTGCTC  
1051 CGCTAA

This corresponds to the amino acid sequence <SEQ ID 446; ORF111>:

5           1   MPSETRLPNF IRVLIFALGF IFLNACSEQT AQTVTLQGET MGTTYTVKYL  
          51   SNNRDKLPSP AEIQKRIDDA LKEVNRQMS YQPDSEISRF NQHTAGKPLR  
         101   ISSDFAHVTA EAVRLNRLTH GALDVTVGPL VNLWGFDPK SVTREPSPEQ  
         151   IKQAASYTGI DKIIKQKGD YASLSKTHPK AYLDLSSIAK GFGVDKVAGE  
         201   LEKYGIQNYL VEIGGELHGK GKNARGEPRW IGIEQPNIVQ GGNTQIIVPL  
         251   NNRSLATSGD YRIFHVDKNG KRLSHIINPN NKRPISHNLA SISVVADSAM  
10           301   TADGLSTGLF VLGETEALKL AEREKLAVFL IVRDKGGYRT AMSSEFEKLL  
         351   R\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF111 shows 96.9% identity over a 351aa overlap with an ORF (ORF111a) from strain A of *N.*

15 *meningitidis*:

		10	20	30	40	50	60
	orf111a.pep	MPSETRLPNFIRTLIFALSFIFLNACSEQT AQTVTLQGETMGTTYTVKYLSNNRDXLPSP					
	orf111	:     :     :     :     :     :     :					
20		10	20	30	40	50	60
	orf111a.pep	AEIQKRIDDALKEVNRQMSYQPDSEISRFNQHTAGKPLRISSDFAHVTA EAVHLNRLTH					
	orf111	:     :     :     :     :     :     :					
25		70	80	90	100	110	120
	orf111a.pep	AEIQKRIDDALKEVNRQMSYQPDSEISRFNQHTAGKPLRISSDFAHVTA EAVHLNRLTH					
	orf111	:     :     :     :     :     :     :					
		70	80	90	100	110	120
30		130	140	150	160	170	180
	orf111a.pep	GALDVTVGPLVNLWGFDPKSVTREPSPEQIKQAASYTGIDKIIKQKGDYASLSKTHPK					
	orf111	:     :     :     :     :     :     :					
		130	140	150	160	170	180
35		190	200	210	220	230	240
	orf111a.pep	AYLDLSSIAKGFGVDXVAGELEKYGIQNYLVEIGGELHGKXKNARGEPRWIGIEQPNIVQ					
	orf111	:     :     :     :     :     :     :					
		190	200	210	220	230	240
40		250	260	270	280	290	300
	orf111a.pep	GGNTQIIVPLNNRSXATSGDYRIFHVDKNGKRLSHIINPNKRPISHNLASISVXADSAM					
	orf111	:     :     :     :     :     :     :					
		250	260	270	280	290	300
45		310	320	330	340	350	
	orf111a.pep	TADGXSTGLFVLGETEALKLAEREKLAVFLIVRDKGGYRTAMSSEFEKLLRX					
	orf111	:     :     :     :     :     :     :					
50		310	320	330	340	350	

The complete length ORF111a nucleotide sequence <SEQ ID 447> is:

55           1   ATGCCGTCTG AAACACGCCT GCCGAACCTT ATCCGCACCT TGATATTTGC  
          51   CCTGAGTTTT ATCTTCCTGA ACGCTGTTC GGAACAAACC GCGCAAACCG  
         101   TTACCCTGCA AGGTGAAACG ATGGGCACGA CCTATACCGT CAAATACCTT  
         151   TCAAATAATC GGGACNAACT CCCNTCACCT GCCGAAATAC AAAANCGCAT  
         201   CGATGACGCG CTAAAGAAG TCAACCGGCA GATGTCCACC TATCAGCCCG  
         251   ACTCCGAAAT CAGCCGGTTC AACCAACACA CAGCCGGCAA GCCCCTCCGC  
         301   ATTTCAAGCG ACTTCGCACA CGTTACTGCC GAAGCCGTCC ACCTGAACCG  
         351   CCTGACACAC GCGCGCTGG ACGTAACCGT CGGCCCTTG GTCAACCTT  
60           401   GGGGATTTCG CCCCACAAA TCCGTTACCC GTGAACCGTC GCCGGAACAA

5

10

15

20

### Homology with a predicted ORF from *N.gonorrhoeae*

25 ORF111 shows 96.6% identity over a 351aa overlap with a predicted ORF (ORF111.ng) from *N. gonorrhoeae*:

40

45

50

55

60

The complete length ORF111ng nucleotide sequence <SEQ ID 449> is:

1 ATGCCGCTCTG AAACACGCCT GCCGAACCTT ATCCGCGCCT TGATATTTGC  
 51 CCTGGGTTTC ATCTTCCTGA ACGCCTGTTC GGAacaaacC GCGCAaaccg  
 101 TTACCCTGCA AGGCGAAAcg aTGGGTACGA CCTATACCGT CAAATACCTT  
 151 TCAAATAATC GGGACAAACT CCCCTCCCCT GCCAAAATAC AAAAGCGCAT  
 201 TGATGATGCG CTAAAGAAG TCAACCGGCA GATGTCCACC TACCAGACCG  
 251 ATTCCGAAAT CAGCCGGTTC AACCAACACA CAGCCGCGAA GCCCCTCCGC  
 301 ATTTCAAGCG ATTTCCGACA CGTTACCGCC GAAGCCGTCC GCCTGAACCG  
 351 CCTGACTCAC GGCGCACTGG ACGTAACCGT CGGCCCTTTG GTCAACCTTT  
 401 GGGGGTTCGG CCCCACAAA TCCGTTACCC GTGAACCGTC GCCGGAACAA  
 451 ATCAAACAGG CGGCATCTTA TACGGGCATA GACAAAATCA TTTTGCAACA  
 501 AGGCAAAGAT TACGCTTCTT TGAGCAAAAC CCACCCAAA GCCTATTGG  
 551 ATTTATCTTC GATTGCCAAA GGCTTCGGCG TTGATAAAGT TCGGGGCGAA  
 601 CTGGA AAAAT ACGGCATTCA AAATTATCTG GTCGAAAtcg gcggcGAGTT  
 651 GCACGGCAAA GGCAAAAATG CGCACGGCGA ACCGTGGCGC ATCGGTATAG  
 701 AGCAACCCAA TATCATCCAA GcgGCAata CGCAGATTat cgtcccgctg  
 751 aaCaaccgtt cgtTGCCAC TTCCGGCGAT TAccgtaTTT tccacgtcgA  
 801 TAAAAcggc aaacgccttt cccacaTCAT CAATCCCaAC aacAAACgac  
 851 ccATCAGcca caacctcgcc tccatcagcg tggctctcAGA CAGTGCATG  
 901 ACGGCGGACG GTTtatCCAC AGGATTATTT GTTTTAGCG AAACCGAAGC  
 951 CTTAAGGCTG GCAGAACAAG AAAAActCGC TGTTTTCCTA ATTGTCCGGG  
 1001 ATAAGGACGG CTACCGCACC GCCATGTCTT CCGAATTGTC CAAGCTGCTC  
 1051 CGCTAA

This encodes a protein having amino acid sequence <SEQ ID 450>:

1 MPSETRLPNL IRLALIFALGF IFLNACSEQT AQTVTTLQGET MGTITYTVKYL  
 51 SNNRDKLPSP AKIQKRIDDA LKEVNRQMST YQTDSEISRF NQHTAGKPLR  
 101 ISSDFAHVTA EAVRLNRLTH GALDVTGVP LNLWGFPGDK SVTREPSPEQ  
 151 IKQAASYTGI DKILQQGKD YASLSKTHPK AYLDLSSIAK GFGVDKVAGE  
 201 LEKYGIQNYL VEIGGELHGK GKNAHGEPRW IGIEQPNIIQ GNTQIIIVPL  
 251 NNRSLSATSGD YRIFHVDKNG KRLSHIINPN NKRPISHNLA SISVVSASAM  
 301 TADGLSTGLF VLGETEALRL AEQEKLAFLV IVRDKDGYRT AMSSEFAKLL  
 351 R\*

This protein shows homology with a hypothetical lipoprotein precursor from *H. influenzae*:

sp|P44550|YOJL HAEIN HYPOTHETICAL LIPOPROTEIN HI0172 PRECURSOR >gi|1074292|pir|4  
 hypothetical protein HI0172 - Haemophilus influenzae (strain Rd KW20)  
 >gi|1573128 (U32702) hypothetical [Haemophilus influenzae] Length = 346  
 Score = 353 bits (896), Expect = 9e-97  
 Identities = 181/344 (52%), Positives = 247/344 (71%), Gaps = 4/344 (1%)  
 Query: 7 LPNLIRLALIFALGFIFLNACSEQTAQTVTTLQGETMGTITYTVKYLNNRDKLPSPAKIQKR 66  
 + LI +I + L AC ++T + ++L G+TMGTTY VKYL + S K +  
 Sbjct: 1 MKKLISGIIAVAMALSLAACQKET-KVISLSGKTMGTITYHVKYLDGSIATASE-KTHEE 58  
 Query: 67 IDDALKEVNRQMSTYQTDSEISRFNQHT-AGKPLRISSDFAHVTAEAVRLNRLTHGALDV 125  
 I+ LK+VN +MSTY+ DSE+SRFNQ+T P+ IS+DFA V AEA+RLN++T GALDV  
 Sbjct: 59 IEAILKDVNAKMSTYKDSLSRFNQHTQVNTPIEISADFAKVLAEAIRLNKVTGALDV 118  
 Query: 126 TVGPLVNLWGFPGDKSVTREPSPEQIKQAASYTGIDKILQQGKDYASLSKTHPKAYLDL 185  
 TVGP+VNLWGFPG+K ++P+PEQ+ + ++ GIDKI L K+ A+LSK P+ Y+DL  
 Sbjct: 119 TVGPVNLWGFPGPEKRPEKQPTPEQLAERQAWVGIDKITLDTNKEKATLSKALPQVYVDL 178  
 Query: 186 SSIAGFGVDKVAGELEKYGIQNYLVEIGGELHGKGNHGEPRWIGIEQPNIIQGGNTQ 245  
 SSIAGFGVD+VA +LE+ QNY+VEIGGE+ KGKN G+PW+I IE+P +  
 Sbjct: 179 SSIAGFGVDQVAEKLEQLNAQNYMVEIGGEIRAKGKNIEGKWPQIAIEKPTTTGERAVE 238  
 Query: 246 IIVPLNNRSLSATSGDYRIFHVDKNGKRLSHIINPNKRPISHNLASISVVSASAMTADGL 305  
 ++ LNN +A+SGDYRI+ ++NGKR +H I+P PI H+LASI+V++ ++MTADGL  
 Sbjct: 239 AVIGLNNMGMASSGDYRIY-FEENGKRFHEIDPKTGYPYIQQHSLASITVLAPTSMTADGL 297  
 Query: 306 STGLFVLGETEALRLAEQEKLAFLVLRDKDGYRTAMSSEFAKL 349  
 STGLFVLGE +AL +AE+ LAV+LI+R +G+ T SS F KL  
 Sbjct: 298 STGLFVLGEDKALEVAEKNLAVYLIIRTDNGFVTKSSSAFKKL 341

Based on this analysis, it is predicted that the proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

**Example 54**

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 451>:

```

1  ..CCGTGCCGCC GACAGGGCGA CGACGTGTAT GCGGCGCACG CGTCCCCTCA
51 AAAATTGTGG CTGCGCTTCA TCGGCGGCCG GTCGCATCAA AATATACGGG
101 GCGGCGCGGC TCGGACGGG TGGCGCAAAG GCGTGCAAAT CCGGCGGCGAG
151 GTGTTGTAC GGCAAAATGA AGGCAGCCKA yTGGCAATCG GCGTGATGGG
201 CGGCAGGGCC GGCCAGCACG CwTCAGTCAA CGGCAAAGGC GGTGCGGCAG
251 gCAGTGATTT GTATGTTAT GgCGGGGgTG TTTATGCTgC GTGGCATCAG
301 TTGCGCGATA AACAAACGGG TgCGTATTTG GACGGCTGGT TGCAATACCA
10 351 ACGTTTCAAA CACCGCATCA ATGATGAAAA CCGTGCGGAA CgCTACAAAA
401 CCAAAGGTTG GACGGCTTCT GTCGAAGGCG GCTACAACGC GCTTGTGGCG
451 GAAGGCATTG TCGGAAAAGG CAATAATGTG CCGTTTACC TACAACCGCA
501 GgCGCAGTTT ACCTACTTGG GCGTAAACGG CCGCTTACC GACAGCGAGG
551 GGACGCGCGT CCGACTGCTC GGACGCGGTC AGTGGCAAAG CCGCGCCGGC
15 601 AtTCGGGCAA AAACCGGTTT TGCTTTGCGT AACGCTGTCA ATCTTCAGCC
651 TTTTGGCGCT TTTAATGttT TGCACAGGTC AAAATCTTTC GCGGTGAAAA
701 TGGACGCGCA AAAACAGACG CTGGCAGGCA GGACGGCACT CGAAGGGCGG
751 TTCGGTATTG AAGCCGGTTG GAAAGGCCAT ATGTCCGCA..

```

This corresponds to the amino acid sequence <SEQ ID 452; ORF35>:

```

20 1  ..PCRRQGDDVY AAHASRQKLW LRFIGGRSHQ NIRGGAAADG WRKGVQIGGE
51 VFVRQNEGSX LAIGVMGGRA GQHASVNGKG GAAGSDLYGY GGGVYAAWHQ
101 LRDKQTGAYL DGWLQYQRFK HRINDENRAE RYKTKGWAS VEGGYNALVA
151 EGIVGKGNNV REYLPQQAQF TYLGVNGGFT DSEGTAVGLL GSGQWQSRAG
201 IRAKTRFALR NGVNLQPFPA FNVLHRSKSF GVEMDGEKQT LAGRTALEGR
25 251 FGIEAGWKGH MSA..

```

Computer analysis of this amino acid sequence gave the following results:

Homology with putative secreted VirG-homologue of *N. meningitidis* (accession number A32247)

ORF and virg-h protein show 51% aa identity in 261aa overlap:

```

30 Orf35 5 QGDDVYAAHASRQKLWLRFIGGRSHQNIRGGAA-ADGWRKGVQIGGEVFVRQNEGSXLAI 63
+ D++ R+ LWLR I G S+Q ++G A +G+RKGVQ+GGEVF QNE + L+I
virg-h 396 KNSDIFDRTLPRKGLWLRVIDGHSNQVQKGTAPVEGYRKGVQLGGEVFTWQNESNQLSI 455

35 Orf35 64 GVMGGRAGQHASVNGKG--GAAGSDLYGYGGGVYAAWHQLRDKQTGAYLDGWLQYQRFKH 121
G+MGG+A Q ++ + ++ G+G GVYA WHQL+DKQTGAY D W+QYQRF+H
virg-h 456 GLMGQAQERSTFHNPDNDLTTGNVKGFGAGVYATWHQLQDKQTGAYADSWMQYQRFH 515

40 Orf35 122 RINDENRAERYKTKGWASVEGGYNALVAEGIVGKGNNVREYLPQQAQFTYLGVNGGFTD 181
RIN E+ ER+ +KG TAS+E GYNAL+AE KGN++R YLQPPAQ TYLGVNG F+D
virg-h 516 RINTEDGTERFTSKGITASIEAGYNALLAEHFTKKNLSRVLYLPQAQLTYLGVNGKFSD 575

45 Orf35 182 SEGTAVGLLGSGQWQSRAGIRAKTRFALRNGVNLQPFPAFNVLHRSKSFGEVMDGEKQTL 241
SE V LLGS Q Q+R G++AK +F+L + ++PFAA N L+ +K FGVEVDGE++ +
virg-h 576 SENAHVNLLGSRQLQTRVGVQAKAQFSLYKNIAIEPFAAVNALYHNKPFGEVMDGERRVI 635

50 Orf35 242 AGRTALEGRFGIEAGWKGHMS 262
+TA+E + G+ K H++
virg-h 636 NNKTAIESQLGVAVKIKSHLT 656

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF35 shows 96.9% identity over a 259aa overlap with an ORF (ORF35a) from strain A of *N. meningitidis*:

```

55 orf35.pep
10 20 30
PCRRQGDDVYAAHASRQKLWLRFIGGRSHQNIRG
:||||| |

```

orf35a QRLAIPAEAVLYAQQAANTLFLGLRAADRDDVYAADPSRQKLWLRFIGGRSHQNIRG  
310 320 330 340 350 360

5 orf35.pep 40 50 60 70 80 90  
GAAADGWRKGVQIGGEVFRQNEGSXLAIGVMGGRAGQHASVNGKGGAAGSDLYGYGGGV  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| :|||||  
orf35a GAAADGRRKGVQIGGEVFRQNEGSRLAIGVMGGRAGQHASVNGKGGAAGSYLHGYGGGV  
370 380 390 400 410 420

10 orf35.pep 100 110 120 130 140 150  
YAAWHQLRDKQTGAYLDGWLQYQRFKHRINDENRAERYKTKGWTASVEGGYNALVAEGIV  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| :|||||  
orf35a YAAWHQLRDKQTGAYLDGWLQYQRFKHRINDENRAERYKTKGWTASVEGGYNALVAEGVV  
430 440 450 460 470 480

15 orf35.pep 160 170 180 190 200 210  
GKGNNVRFYLPQAQFTYLGVNGGFTDSEGTAVGLLGSQWQSRAGIRAKTRFALRNGVN  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| :|||||  
orf35a GKGNNVRFYLPQAQFTYLGVNGGFTDSEGTAVGLLGSQWQSRAGIRAKTRFALRNGVN  
490 500 510 520 530 540

20 orf35.pep 220 230 240 250 260  
LQPFAAFNVLHRSKSFVEMDGEKQTLAGRTALEGRFGIEAGWKGHMSA  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| :|||||  
orf35a LQPFAAFNVLHRSKSFVEMDGEKQTLAGRTALEGRFGIEAGWKGHMSARIGYGKRTDGD  
550 560 570 580 590 600

25 orf35a KEAALSILKWLFX  
610 620

30 The complete length ORF35a nucleotide sequence <SEQ ID 453> is:

1 ATGTTTCAGAG CTCAGCTTGG TTCAAATACT CGTTCTACCA AAATCGGCGA  
51 CGATGCCGAT TTTTCATTTT CAGACAAGCC GAAACCCGGC ACTTCCCATT  
101 ATTTTTCAG CGGTAAAACC GATCAAAAT CATCCGAATA TGGGTATGAC  
151 GAAATCAATA TCCAAGGTAA AACTACAAT AGCGGCATAC TCGCGTCGA  
201 TAATATGCCC GTTGTAAAGA AATATATTAC AGATACTTAC GGGGATAATT  
251 TAAAGGATGC GGTAAAGAAG CAATTACAGG ATTTATACAA AACAAGACCC  
301 GAAGCTTGGG AAGAAAATAA AAAACGGACT GAGGAGGCGT ATATAGAACA  
351 GCTTGGACCA AAATTTAGTA TACTCAAACA GAAAACCCC GATTTAATTA  
401 ATAAATTGGT AGAAGATTCC GTACTCACTC CTCATAGTAA TACATCACAG  
451 ACTAGTCTCA ACAACATCTT CAATAAAAAA TTACACGTCA AAATCGAAAA  
501 CAAATCCAC GTCGCCGAC AGGTGTTGGA ACTGACCAAG ATGACGCTGA  
551 AAGATTCCCT TTGGGAACCG CGCCGCCATT CCGACATCCA TATGCTGGAA  
601 ACTCCGATA ATGCCGCAT CCGCTGAAC ACGAAAGATG AAAAAGTAC  
651 CGTCCATAAA GCGTATCAGG GCGGTGCGGA TTTCCTGTTC GGCTACGACG  
701 TGCGGGAGTC GGACAAACCC GCCCTGACCT TTGAAGAAAA AGTCAGCGGA  
751 CAATCCGGCG TGGTTTGGGA ACGCCGGCCG GAAAATCTGA AAACGCTCGA  
801 CGGGCGCAAA CTGATTGCGG CGGAAAAGGC AGACTCTAAT TCGTTGCGT  
851 TTAAACAAAA TTACCGGCAG GACTGTACG AATTATTGCT CAAGCAATGC  
901 GAAGGCGGAT TTTGCTTGGG CGTGCAGCGT TTGGCTATCC CCGAGGCGGA  
951 AGCGGTTTTA TATGCCAAC AGGCTTATGC GGCAAATACT TTGTTCGGGC  
1001 TGCGTGCCGC CGACAGGGGC GACGACGTGT ATGCCGCCGA TCCGTCCCCT  
1051 CAAAATTGT GGCTGCGCTT CATCGCGCGC CGGTGCGATC AAAATATACG  
1101 GGGCGGCGCG GCTGCGGACG GCGGCGCAA AGGCGTGCAA ATCGGCGGCG  
1151 AGGTGTTTGT ACGGCAAAAT GAAGGCAGCC GGCTGGCAAT CGGCGTGATG  
1201 GCGGCGAGGG CTGGCCAGCA CGCATCAGT AACGGCAAAG GCGGTGCGGC  
1251 AGGCAGTTAT TTGCATGGTT ATGGCGGGGG GTTTTATGCT GCGTGGCATC  
1301 AGTTGCGCGA TAAACAAACG GGTGCGTATT TGGACGGCTG GTTGCAATAC  
1351 CAACGTTTCA AACACCGCAT CAATGATGAA AACCGTGGG AACGCTACAA  
1401 AACCAAAGGT TGGACGGCTT CTGTGGAAG CCGTACAAC GCGTGTGTG  
1451 CGGAAGGCGT TGTGCGAAAA GGCAATAATG TGCGGTTTTA CCTGCAACCG  
1501 CAGGCGCAGT TTACTTACTT GGGCGTAAAC GCGGCTTTTA CCGACAGCGA  
1551 GGGGACGGCG GTCGGAAGTC TCGGCGAGCG TCAGTGGCAA AGCCGCGCCG  
1601 GCATTGCGGC AAAAACCCTT TTTGCTTTGC GTAAACGGTGT CAATCTTCAG  
1651 CCTTTTGCCG CTTTAAATGT TTTGCACAGG TCAAATCTT TCGGCGTGGA  
1701 AATGGACGCG GAAAAACAGA CGCTGGCAGG CAGGACGGCG CTCGAAGGCG  
1751 GGTTCGGCAT TGAAGCCGGT TGGAAAGGCC ATATGTCCGC ACGCATCGGA  
1801 TACGGCAAAA GGACGGACGG CGACAAAGAA GCCGCATTGT CGCTCAAATG  
1851 GCTGTTTTGA

This encodes a protein having amino acid sequence <SEQ ID 454>:

```

1 MFRAQLGSNT RSTKIGDDAD FSFSDKPKPG TSHYFSSGKT DQNSSEYGYD
51 EINIQQKNYN SGILAVDNMP VVKKIYITDY GDNLKDAVKK QLQDLYKTRP
101 EAWEEKKRT EEAYIEQLGP KFSILKQKNP DLINKLVEDS VLTPHSNTSQ
5 151 TSLNNIFNKK LHVKIENKSH VAGQVLELTK MTLKDSLWEP RRHSDIHMLE
201 TSDNARIRLN TKDEKLTVHK AYQGGADFLF GYDVRESDKP ALTFEEKVSG
251 QSGVVLERRP ENLKTLDGRK LIAAEKADSN SFAFKQNYRQ GLYELLKQOC
301 EGGFCLGVQR LAIPEAEAVL YAQQAYAANT LFGLRAADRG DDVYAADPSR
351 QKLWLRFIGG RSHQNIIRGGA AADGRRKGVQ IGGEVFVRQN EGSRLAIGVM
10 401 GGRAGQHASV NGKGAAGSY LHGYGGGVYA AWHQLRDKQT GAYLDGWLQY
451 QRFKHRINDE NRAERYKTKG WTASVEGGYN ALVAEGVVGK GNNVRFYLPQ
501 QAQFTYLGVN GGFTDSEGTA VGLLGSGQWQ SRAGIRAKTR FALRNGVNLQ
551 PFAAFNVLHR SKSFGVEMDG EKQTLAGRTA LEGRFGIEAG WKGHMSARIG
601 YGKRTDGDKE AALSLKWL*

```

#### 15 Homology with a predicted ORF from *N.gonorrhoeae*

ORF35 shows 51.7% identity over a 261aa overlap with a predicted ORF (ORF35ngh) from *N.gonorrhoeae*:

```

20 orf35.pep          PCRRQGGDDVYAAHASRQKLWLRFIGGRSHQNIIRG 34
   :::::  ::  |||  |  |::|  ::|
orf35ngh          FTKVQERDDIAYAAQAQAANTLFALRLNDKNSDIFDRTLPRKGLWLRVIDGHSNQWVQG 370
orf35.pep          GAA-ADGWRKGVQIGGEVFVRQNEGSXLAIGVMGGRAGQHASVNGK--GAAGSDLYGYG 91
   :|  ::|:|||||:|||||: ||::|  |::|:|||||  |:::  :  :  :  :  :|
orf35ngh          KTAPVEGYRKGVQLGGEVFTWQNESNQLSIGLMGGQAEQRSTFRNPDTNLTGTVNGKFG 430
25 orf35.pep          GGVYAAWHQLRDKQTGAYLDGWLQYQRFKHRINDENRAERYKTKGWTASVEGGYNALVAE 151
   :||||:||||:||||:||||:||||:||||  |  :||:  :||  |||:||||:|
orf35ngh          AGVYATWHQLQDKQTGAYVDSWMQYQRFHRINTEYATERFTSKGITASIEAGYNALLAE 490
30 orf35.pep          GIVGKGNVRFYLPQAQFTYLGVNNGFTDSEGTA VGLLGSGQWQSRAGIRAKTRFALRN 211
   ::  |||::|  |||||:|||||  |::|:  |::|  |  |||:|:|:|:|:|
orf35ngh          HFTKKGNSLRVYLPQAQLTLYLVNGKFSDSENAQVNLLGSRQLQSRVGVQAKAQFAFTN 550
35 orf35.pep          GVNLPFAAFNVLHRSKSGVEMDGEKQTLAGRTALEGRFGIEAGWKGHMSA 263
   ||::|||:|  |  ::::|  |||:|:|:|:  :|::|  :|:  |  |:|:
orf35ngh          GVTFPQPFVAVNSIYQQKPFGEIDGDRRVINNKTVIETQLGVAAKIKSHLTQASFNRQT 610

```

A partial ORF35ngh nucleotide sequence <SEQ ID 455> is predicted to encode a protein having partial amino acid sequence <SEQ ID 456>:

```

40 1 . .KKLRDRNSEY WKEETYHIKS NGRTPNIPA LFPKHPFDPF ENINNSKKIS
51 FYDKEYTEDY LVGFARGFGV EKRNGEEKP LRQYFKDCVN TENSNDNCK
101 ISSFGNYGPI LIKSDIFALA SQIKNSHINS EILSVGNYIE WLRPTLNKLT
151 GWQEHLYAGL DPFHYIEVTD NSHVIGQID LGALELTNSL WKPRWNSNID
201 YLITKNAEIR FNTKNESLLV KEDYAGGARF RFAYDLKDKV PEIPVLTFEK
45 251 NITGTSDIIF EGKALDNLKH LDGHQIVKVN DTADKDAFRL SSKYRKGIYT
301 LSLQQRPEGF FTKVQERDDI AIYAAQAQAA NTLFALRLND KNSDIFDRTL
351 PRKGLWLRVI DGHSNQWVQG KTAPVEGYRK GVQLGGEVFT WQNESNQLSI
401 GLMGGQAEQR STFRNPDTN LTTGNVKGFG AGVYATWHQL QDKQTGAYVD
451 SWMQYQRFH RINTEYATER FTSKGITASI EAGYNALLAE HFTKKGNSLR
50 501 VYLQPPAQLT YLGVNGKFS SENAQVNLLG SRQLQSRVGV QAKAQFAFTN
551 GVTFPQPFVAV NSIYQQKPFGE VEIDGDRVI NNKTVIETQL GVAKIKSHL
601 TLQASFNRQT SKHHHAKQGA LNLQWTF*

```

Based on this prediction, these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 55

55 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 457>:

5  
10  
15  
20  
25  
30  
35

```

1  ..GCGGAATATG TTCAGTTCTC TATAGATTG TTCAGTGTGG GTAAATCGGG
51  GGGCGGTATA CCTAAGGCTA AGCCTGTGTT TGATGCGAAA CCGAGATGGG
101 AGGTTGATAG GAAGCTTAAT AAATTGACAA CTCGTGAGCA GGTGGAGAAA
151 AATGTTTCAGG AAACGAGAAG AAGGAGTCAG AGTAGTCAGT TTAAAGCCCA
201 TGCGCAACGA GAATGGGAAA ATAAACAGG GTTAGATTTT AATCATTTTA
251 TAGGTGGTGA TATCAATAAA AAAGGCACAG TAACAGGAGG GCATAGTCTA
301 ACCCGTGGTG ATGTACGGGT GATACAACAA ACCTCGGCAC CTGATAAACA
351 TGGGGT.TTA TCAAGCGACA GTGGAATTN A

```

This corresponds to the amino acid sequence <SEQ ID 458; ORF46>:

10  
15  
20

```

1  ..AEYVQFSIDL FSVGKSGGGI PKAKPVFDAK PRWEVDRKLN KLTTRREQVEK
51  NVQETRRRSQ SSQFKAHAQR EWENKTGLDF NHFIGGDINK KGTVTGGHSL
101 TRGDVRVIQQ TSAPDKHGXL SSDSGNX

```

Further work revealed further partial nucleotide sequence <SEQ ID 459>:

15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65

```

1  ..GCAGTGTGCC TnCCGATGCA TGCACACGCC TCAAnATTGG CAAACGATTC
51  TTTTATCCGG CAGGTTCTCG ACCGTGAGCA TTTCGAACCC GACGGGAAAT
101 ACCACCTATT CGGCAGCAGG GGGGAACCTG CCGAGCGCCA GTCTCATATC
151 GGATTGGGAA AAATACAAAG CCATCAGTTG GGCAACCTGA TGATTCAACA
201 GGCGCCATT AAAGGAAATA TCGGCTACAT TGTCCGCTTT TCCGATCACG
251 GGCACGAAGT CCATTCCCCs TTCGACAACC ATGCCTCACA TTCCGATTCT
301 GATGAAGCCG GTAGTCCCGT TGACGGATTG AGCCTTTACC GCATCCATTG
351 GGACGGATAC GAACACCATC CCGCCGACGG CTATGACGGG CCACAGGGCG
401 GCGGCTATCC CGCTCCCAAA GCGCGGAGGG ATATATACAG TTACGACATA
451 AAAGGCGTTG CCCAAATAT CCGCCTCAAC CTGACCGACA ACCGAGCAC
501 CGGACAACGG CTGTCCCGACC GTTCCACAA TGCCGGTAGT ATGCTGACGC
551 AAGGAGTAGG CGACGGATTG AAACGCGCCA CCCGATACAG CCCCAGAGCTG
601 GACAGATCGG GCAATGCCCG CGAAGCCTTC AACGGCACTG CAGATATCGT
651 TAAAAACATC ATCGGCGCTG CAGGAGAAAT TGT

```

This corresponds to the amino acid sequence <SEQ ID 460; ORF46-1>:

30  
35  
40  
45  
50

```

1  ..AVCLPMHAHA SXLANDSFIR QVLDROHFEP DGKYHLFGSR GELAERQSHI
51  GLGKIQSHQL GNLMIQQAII KGNIGYIVRF SDHGHEVHSP FDNHASHSDS
101 DEAGSPVDGF SLYRIHWDGY EHHPADGYDG PQGGGYPAK GARDIYSYDI
151 KGVAQNIRLN LTDNRSTGQR LADRFRHNAGS MLTQGVGDGF KRATRYSPLE
201 DRSGNAEAF NGTADIVKNI IGAAGEI

```

Computer analysis of this amino acid sequence gave the following results:

### 35 Homology with a predicted ORF from *N.gonorrhoeae*

ORF46 shows 98.2% identity over a 111aa overlap with a predicted ORF (ORF46ng) from *N.gonorrhoeae*:

40  
45

```

orf46.pep          AEYVQFSIDLF SVGKSGGGI PKAKPVFDAK PRWEVDRKLNKLTR 45
                    |||
orf46ng            PKTGVPFDGKGFPNFEKHVKYDTKLDIQELSGGGI PKAKPVFDAK PRWEVDRKLNKLTR 217
                    |||
orf46.pep          EQVEKNVQETRRRSQSSQFKAHAQREWENKTGLDFNHFIGGDINKKGTVTGGHSLTRGDV 105
                    |||
orf46ng            EQVEKNVQETRRRSQSSQFKAHAQREWENKTGLDFNHFIGGDINKKGA VTGGHSLTRGDV 277
                    |||
orf46.pep          RVIQQTSA PDKHGXLSSDSGN 126
                    |||
orf46ng            RVIQQTSA PDKHGVLSSDSGN 298

```

A partial ORF46ng nucleotide sequence <SEQ ID 461> is predicted to encode a protein having

50 partial amino acid sequence <SEQ ID 462>:

```

1  ..RRLKHCCHAR LGSFHRKQD GAHQRFGRYG ATQRLCRSSH PRLGSPKPOC
51  RTRHRSRQQY LYGSHPHQRD WSCPGKIQLG RHGHTSCRAV ADXRDRICER
101 EIRRQRQXCR CRLGKIPSLs IPKYPLKLEQ RYKGENITSS TVPPSNGKNV
151 KLADQRHPKT GVPFDGKGFP NFEKHVKYDT KLDIQELSGG GIPKAKPVFD

```



201 AKPRWEVDRK LNKLTTRQV EKNVQETRRR SQSSQFKAHA QREWENKTGL  
251 DENHFIGGDI NKKGAVTGGH SLTRGDVIRVI QQTSAPDKHG VLSSDSGN\*

Further work revealed the complete gonococcal DNA sequence <SEQ ID 463>:

5 1 TTGGGCATT CCCGCAAAAT ATCCCTTATT CTGTCCATAC TGGCAGTGTG  
51 CCTGCCGATG CATGCACACG CCTCAGATTT GGcaAACGAT CCCTTTATCC  
101 GgCaggtttC CGaccGTcAG CATTTTCgaac ccgacggGaa ATACCaCCTA  
151 TTcggCaGCA GGGGGGAGCT TgcnagcGC aacgcccATa tcggattggG  
201 aaacaTAcAa Agccatcagt tGggccacct gatgattcaa caggcggccg  
251 ttgaaggaaA TAtcgGctac attgtccgct tttccgatca cgggcacaaa  
10 301 ttccattcgc cettcGAcaa ceaTGCCTCA CATTCCGATT CTGACGAAGC  
351 CGGTAGTCCC GTTGACGGAT TCAGCCTTTA CCGCATCCAT TGGGACGGAT  
401 ACGAACACCA TCCCGCCGAC GGCTATGACG GGCCACAGGG CGGCGGCTAT  
451 CCCGCTCCCA AAGGCGCGAG GGATATATAC AGCTACGACA TAAAGGCGT  
501 TGCCCAAAAT ATCCGCCTCA ACCTGACCGA CAACCCGAGC ACCGACCAAC  
15 551 GGCTTGCCGA CCGTTTCCAC AATGCCGCG CTATGCTGAC GCAAGGAGTA  
601 GGGACGCGAT TCAAACGCGC CACCCGATAC AGCCCCGAGC TGGCAGATC  
651 GGGCAATGcc gccGAAGCCT TCAACGGCAC TGCAGATATC GTCAAAAACA  
701 TCATCGGCGC GGCAGGAGAA ATGTCTGGCG CAGGCGATGC CGTGcagGGT  
20 751 ATAAGCGAAG GCTCAAACAT TGCTGTcATG CACGGCTTGG GTCTGCTTTC  
801 CACCGAAAAC AAGATGGCGC GCATCAACGA TTTGGCAGAT ATGGCGCAAC  
851 TCAAAGACTA TGCCGcAGCA GCCATCCGCG ATTGGGCAGT CCAAAACCCC  
901 AATGCCGCAC AAGGCATAGA AGCCGTcAGC AATATCTTTA TGGCAGCCAT  
951 CCCCATCAAA GGgATTGGAG CTGTCCGGGG AAAATACGGC TTGGGCGGCA  
25 1001 TCACGGCACA TCCTGTCAAG CGGTcGCAGA TGGGCGCGAT CGCATGCGG  
1051 AAAGGGAAT CCGCCGTcAG CGACAATTT GCCGATGCGG CAtACGCCAA  
1101 ATACCCGTCC CCTTACCATT CCCGAAATAT CCGTTCAAAC TTGGAGCAGC  
1151 GTTACGGCAA AGAAAACATC ACCTCCTCAA CCGTGCCGCC GTCAAACGGC  
1201 AAAAATGTCA AACTGGCAGA CCAACGCCAC CCGAAGACAG GCGTACCGTT  
30 1251 TGACGGTAAA GGGTTTCCGA ATTTGAGAA GCACGTGAAA TATGATACGA  
1301 AGCTCGATAT TCAAGAATTA TCGGGGGCG GTATACCTAA GGCTAAGCCT  
1351 GTGTTTGATG CGAAACCGAG ATGGGAGGTT GATAGGAAGC TTAATAAATT  
1401 GACAACTCGT GAGCAGGTGG AGAAAAATGT TCAGGAACG AGAAGAAGGA  
1451 GTCAGAGTAG TCAGTTTAAA GCCATGCGC AACGAGAATG GGAAATAAAA  
35 1501 ACAGGGTTAG ATTTAATCA TTTTATAGGT GGTGATATCA ATAAGAAAGG  
1551 CACAGTAACA GGAGGGCATA GTCTAACCCG TGGTGATGTA CGGGTGATAC  
1601 AACAAACCTC GGCACCTGAT AAACATGGGG TTTATCAAGC GACAGTGGA  
1651 ATTAATAAAGC CTGATGGAAG TTGGGAGGTG AAAACGAAAA AAGGTGGGAA  
1701 AGTGATGACC AAGCACACCA TGTTCCTAAA AGATTGGGAT GAGGCTAGAA  
40 1751 TTAGGGCTGA AGTTACTTCG GCTTGGGAAA GTAGAATAAT GCTTAAGGAT  
1801 AATAAATGGC AGGTACAAG TAAATCGGT ATTAATAATAG AAGGATTTAC  
1851 CGAACCTAAT AGAACAGCAT ATCCCATTTA TGAATAG

This corresponds to the amino acid sequence <SEQ ID 464; ORF46ng-1>:

45 1 LGISRKISLI LSILAVCLPM HAHASDLAND PFIRQVLDRO HFEPDGKYHL  
51 FGSRGELAXR NGHILGNIQ SHQLGHLMIQ QAAVEGNIGY IVRFSDHGK  
101 FHSPFDNHAS HSDSDEAGSP VDGFSLYRIH WDGYEHPAD GYDGPQGGY  
151 PAPKGARDIY SYDIKGVAQN IRLNLTNRS TGQRLADRFH NAGAMLTQGV  
201 GDGFKRATRY SPELDRSGNA AEAENGTAADI VKNIIGAAGE IVGAGDAVQG  
251 ISEGSNIAVM HGLGLLSTEN KMARINDLAD MAQLKDYAAA AIRDWAVQNP  
301 NAAQGIEAVS NIFMAAIIPIK GIGAVRGKYG LGGITAHVPK RSQMGAIALP  
50 351 KGKSAVSDNF ADAAYAKYPS PYHSRNIRSN LEQRYGKENI TSSTVPPSNG  
401 KNVKLADQRH PKTGVPPFDGK GFPNFEKHVK YDTKLDIQEL SGGGIPKAKP  
451 VFDAPRWEV DRKLNKLTR EQVEKNVQET RRRSQSSQFK AHAQREWENK  
501 TGLDENHFIG GDINKKGTVT GGHSLTRGDV RVIQQTSAPD KHGVYQATVE  
55 551 IKKPDGSWEV KTKKGGKVM THTMFPKDWD EARIRAEVTS AWESRIMLKD  
601 NKWQGTSKSG IKIEGFTEPN RTAYPIYE\*

ORF46ng-1 and ORF46-1 show 94.7% identity in 227 aa overlap:

60 orf46-1.pep AVCLPMHAHAXLANDSFIRQVLDROHFEPDGKYHLFGSRGELAER  
orf46ng-1 LGISRKISLILSILAVCLPMHAHASDLANDPFIRQVLDROHFEPDGKYHLFGSRGELAXR  
10 20 30 40 50 60  
50 60 70 80 90 100

orf46-1.pep	QSHIGLGKIQSHQLGNLMIQQAIAKGNIGYIVRFSHDHGHEVHSPFDNHASHSDSDEAGSP
orf46ng-1	NGHIGLGNIQSHQLGHLMIQQAIVEGNIGYIVRFSHDHGKHFSPFDNHASHSDSDEAGSP
	70 80 90 100 110 120
orf46-1.pep	110 120 130 140 150 160
orf46ng-1	VDGFSLYRIHWDGYEHHPADGYDGPQGGGYPAKPGARDIYSYDIKGVAQNIRLNLTNRS
	130 140 150 160 170 180
orf46-1.pep	170 180 190 200 210 220
orf46ng-1	TGQRLADRFHNAGSMLTQGVGDGFKRATRYSPELDRSGNAAEFNGTADIVKNIIGAAGE
	190 200 210 220 230 240
orf46-1.pep	I
orf46ng-1	IVGAGDAVQGISSEGSNIAMHGLGLSTENKMARINDLADMAQLKDYAAAAIRDWAVQNP
	250 260 270 280 290 300

### Homology with a predicted ORF from *N.meningitidis* (strain A)

25 ORF46ng-1 shows 87.4% identity over a 486aa overlap with an ORF (ORF46a) from strain A of *N. meningitidis*:

orf46a.pep	10 20 30 40 50 60
orf46ng-1	LGISRKISILSILAVCLPMHAHASDLANDSFIRQVLDQRHFEPDGKYHLFGSRGELAXR
	10 20 30 40 50 60
orf46a.pep	70 80 90 100 110 120
orf46ng-1	SGHIGLGNIQSHQLGNLFIQQAIAKGNIGYIVRFSHDHGHEVHSPFDNHASHSDSDEAGSP
	70 80 90 100 110 120
orf46a.pep	130 140 150 160 170 180
orf46ng-1	VDGFSLYRIHWDGYEHHPADGYDGPQGGGYPAKPGARDIYSYDIKGVAQNIRLNLTNRS
	130 140 150 160 170 180
orf46a.pep	190 200 210 220 230 240
orf46ng-1	TGQRLADRFHNAGSMLTQGVGDGFKRATRYSPELDRSGNAAEFNGTADIVKNIIGAAGE
	190 200 210 220 230 240
orf46a.pep	250 260 270 280 290 300
orf46ng-1	IVGAGDAVQGISSEGSNIAMHGLGLSTENKMARINDLADMAQLKDYAAAAIRDWAVQNP
	250 260 270 280 290 300
orf46a.pep	310 320 330 340 350 360
orf46ng-1	NAAQIEAVSNIFTAVIPVKIGIAVRGKYGLGGITAHVPKRSQMGEIALPKGKSAVSDNF
	310 320 330 340 350 360
orf46a.pep	370 380 390 400 410 420
orf46ng-1	ADAAAYAKYPSYPHSRNIIRSNLEQRYGKENITSSTVPPSNGKNVLANRHPKTKVPFDGK
	370 380 390 400 410 420

		370	380	390	400	410	420
		430	440	450	460	470	
5	orf46a.pep	GFPNFEKDVKYDTRINTAVPQVN----	PIDEPVFN--	PKGSVGS	SAHSW	SITARIQ	YAKLP
	orf46ng-1	GFPNFEKHVKYDTKLD--	IQELSGGGI	PKAKPVF	DAKPRWEV	DRKLN-K	LTTREQVEKNV
		430	440	450	460	470	
10	orf46a.pep	RQGRIRYIPPKNYS	SPSAPL	PKGPNNG	YLDKFG	NEWT	KGPSRTKGQEF
	orf46ng-1	QETRRRSQSSQ	KAHAQ	REWENKT	GLDFNH	FIGGD	INKKGTVTGGH
		480	490	500	510	520	530

The complete length ORF46a DNA sequence <SEQ ID 465> is:

15	1	TTGGGCATTT	CCGCAAAAT	ATCCCTTATT	CTGTCCATAC	TGGCAGTGTG
	51	CCTGCCGATG	CATGCACACG	CCTCAGATTT	GGCAAACGAT	TCTTTTATCC
	101	GGCAGGTCT	CGACCGTCAG	CATTTCGAAC	CCGACGGGAA	ATACCACCTA
	151	TTCGGCAGCA	GGGGGGA	ACTTGGCAGCGC	AGCGGTCATA	TCCGATTGGG
	201	AAACATACAA	AGCCATCAGT	TGGGCAACCT	GTTTCATCCAG	CAGGCGGCCA
20	251	TTAAAGGAAA	TATCGGCTAC	ATTGTCCGCT	TTTCCGATCA	CGGGCAGCAA
	301	GTCCATTCCC	CCTTCGACAA	CCATGCCTCA	CATTCCGATT	CTGATGAAGC
	351	CGGTAGTCCC	GTTGACGGAT	TCAGCCTTTA	CCGCATCCAT	TGGGACGGAT
	401	ACGAACACCA	TCCCGCCGAC	GGCTATGACG	GGCCACAGGG	CGGCGGCTAT
	451	CCGCTCCCA	AAGGCGCGAG	GGATATATAC	AGTACGACA	TAAAGGCGT
25	501	TGCCCCAAAT	ATCCGCCTCA	ACCTGACCGA	CAACCGCAGC	ACCGGACAAC
	551	GGCTTGTGCGA	CGTTTCCAC	AATACCGGTA	GTATGCTGAC	GCAAGGAGTA
	601	GGCGACGGAT	TCAAACGCGC	CACCCGATAC	AGCCCCGAGC	TGGACAGATC
	651	GGGCAATGCC	GCCGAAGCTT	TCAACGGCAC	TGCAGATATC	GTCAAAAACA
	701	TCATCGGCGC	GGCAGGAGAA	ATTGTCGGCG	CAGGCGATGC	CGTGCAGGGT
30	751	ATAAGCGAAG	GCTCAAACAT	TGCTGTTATG	CACGGCTTGG	GTCTGCTTTC
	801	CACCGAAAAC	AAGATGGCGC	GCATCAACGA	TTTGGCAGAT	ATGGCGCAAC
	851	TCAAAGACTA	TGCCGACGCA	GCCATCCGCG	ATTGGGCAGT	CCAAAACCCC
	901	AATGCCGCAC	AAGGCATAGA	AGCCGTCAGC	AATATCTTTA	CGGCAGTCAT
	951	CCCCGTCAAA	GGGATTTGGAG	CTGTTCCGGG	AAAATACGGC	TTGGGCGGCA
35	1001	TCACGGCACA	TCTGTCAAG	CGTTCGCAGA	TGGGCGAGAT	CGCATTTGCCG
	1051	AAAGGGAAAT	CCGCCGTCAG	CGACAATTTT	GCCGATGCGG	CATACGCCAA
	1101	ATACCCGTCC	CCTTACCATT	CCCGAAATAT	CCGTTCAAAC	TTGGAGCAGC
	1151	GTTACGGCAA	AGAAAACATC	ACCTCCTCAA	CCGTGCCGCC	GTCAAACGGA
40	1201	AAGAATGTGA	AACCTGGCAAA	CAAACGCCAC	CCGAAGACCA	AAGTGCCGTT
	1251	TGACGGTAAA	GGGTTTCCGA	ATTTTGAAAA	AGACGTAAAA	TACGATACGA
	1301	GAATTAATAC	CGCTGTACCA	CAAGTGAATC	CTATAGATGA	ACCCGTCTTT
	1351	AATCCTAAAG	GTTCTGTCCG	ATCGGCTCAT	TCTTGGTCTA	TAAGTCCGAG
	1401	AATTCAATAC	GCAAAATTAC	CAAGGCAAGG	TAGAATCAGA	TATATCCAC
	1451	CTAAAAATTA	CTCTCCTTCA	GCACCGCTAC	CAAAAGGACC	TAATAATGGA
45	1501	TATTTGGATA	AATTTGGTAA	TGAATGGACT	AAAGGTCCAT	CAAGAACTAA
	1551	AGGTCAAGAA	TTTGAATGGG	ATGTTCAATT	GTCTAAAACA	GGAAGAGAGC
	1601	AACTTGGATG	GGCTAGTAGG	GATGGTAAGC	ATTTAAATAT	ATCAATTGAT
	1651	GGAAAGATTA	CACACAAATG	A		

This corresponds to the amino acid sequence <SEQ ID 466>:

50	1	LGISRKISLI	LSILAVCLPM	HAHASDLAND	SFIRQVLDRO	HFEPDGYHL
	51	FGSRGELAER	SGHIGLGNIQ	SHQLGNLFIQ	QAAIKGNIGY	IVRFSDDHGE
	101	VHSPFDNHAS	HSDSDEAGSP	VDGFSLYRIH	WDGYEHHPAD	GYDGPQGGGY
	151	PAPKGARDIY	SYDIKQVAQN	IRLNLTDNRS	TGQRLVDRFH	NTGSMLTQGV
	201	GDGFKRATRY	SPELDRSGNA	AEAFNGTADI	VKNIIIGAAGE	IVGAGDAVQG
55	251	ISEGSNIAMV	HGLGLLSTEN	KMARINDLAD	MAQLKDYAAA	AIRDWAVQNP
	301	NAAQGIEAVS	NIFTAVIPVK	GIGAVRGKYG	LGGITAHVPK	RSQMGEIALP
	351	KGKSAVSDNF	ADAAYAKYPS	PYHSRNIRSN	LEQRYGKENI	TSSTVPPSNG
	401	KNVKLANKRH	PKTKVPFDGK	GFPNFEKDVK	YDTRINTAVP	QVNPIDEPVF
	451	NPKGSVGS	SAHSW	SITARIQY	AKLP	RQGRIRYIPPKNYS
60	501	YLDKFGNEWT	KGPSRTKGQ	EFEWDVQLSKT	GREQLGWASR	DGKHLNISID
	551	SKITHK*				

Based on this analysis, including the presence of a RGD sequence in the gonococcal protein, typical of adhesins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

**Example 56**

5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 467>:

```

1 ATGAATATTC ACACCCTGCT CTCCAAACAA TGGACGCTGC CGCCATTCCCT
51 GCCGAAACGG CTGCTGCTGT CCCTGCTGAT ACTGCTTGCC CCCAATGCGG
101 TGTTTTGGGT TTTGGCACTG CTGACCGCCA CCGCCCGCCC GATTGTCAAT
151 TTGGACTATC TTCCCGCCGC GCTGCTGATC GCCCTGCCTT GGCGTTTCGT
201 CAAAATTGCC GCGGTATTGG CGTTTTGGCT GGCGGTTTTG TTTGACGGGC
251 TGATGATGGT GATCCAACCTC TTCCCTTTTA TGGATCTCAT CGGCGCCATC
301 AACCTCGTCC CCTTCATCCT GACCGCCCCC GCCCCTTATC AGATAATGAC
351 CGGGCTG...
```

This corresponds to the amino acid sequence <SEQ ID 468; ORF48>:

```

15 1 MNIHTLLSKQ WTLPPFLPKR LLLSLILLA PNAVFWVLAL LTATARPIVN
51 LDYLPALLI ALPWRFKIA GVLAFWLAVL FDGLMMVIQL FPFMDLIGAI
101 NLVPFILTAP APYQIMTGL...
```

Further work revealed the complete nucleotide sequence <SEQ ID 469>:

```

20 1 ATGAATATTC ACACCCTGCT CTCCAAACAA TGGACGCTGC CGCCATTCCCT
51 GCCGAAACGG CTGCTGCTGT CCCTGCTGAT ACTGCTTGCC CCCAATGCGG
101 TGTTTTGGGT TTTGGCACTG CTGACCGCCA CCGCCCGCCC GATTGTCAAT
151 TTGGACTATC TTCCCGCCGC GCTGCTGATC GCCCTGCCTT GGCGTTTCGT
201 CAAAATTGCC GCGGTATTGG CGTTTTGGCT GGCGGTTTTG TTTGACGGGC
251 TGATGATGGT GATCCAACCTC TTCCCTTTTA TGGATCTCAT CGGCGCCATC
301 AACCTCGTCC CCTTCATCCT GACCGCCCCC GCCCCTTATC AGATAATGAC
351 CGGGCTGTTG CTGCTGTATA TGCTGGCGAT GCCGTTTGTG TTGCAGAAAG
401 CCGCCGCCAA AACCGACTTC CGGCACATTG CCGTCTGCGC CGCCGTTGTG
451 GCGGCAGCCG GCTATTTCAC CGGCCATTG AGTTACTACG ACCGGGGTCG
501 GATGGCCAAT ATCTTCGGCG CAAACAACCT TACTACGCC AAAAGTCAGG
551 CGATGCTCTA CACGTCAGC CAGAATGCCG ACTTTATTAC CGCCGGCCTG
601 GTCGATCCCG TCTTCCTCCC CTTGGGCAAT CAACAGCGTG CCGCCACGCA
651 TCTGAACGAG CCGAAATCTC AAAAAATCCT CTTTATCGTC GCCGAATCTT
701 GGGGCTGCC GGCCAATCCC GAACTTCAA ACGCCACTTT TGCCAAACTG
751 CTGGCGCAAA AAGACCGTTT TTCGGTTTGG GAAAGCGGCA GTTTTCCCTT
35 801 CATCGGCGCG ACGGTCGAAG GCGAAATGCG CGAAGTGTGT GCCTACGGCG
851 GTTTGCGCGG GTTCGCACTG CGCCGCGCGC CCGACGAAAA ATTTGCCCCG
901 TGCCTCCCCA ACCGTTTGAA ACAAGAAGGT TACGCCACCT TTGCGATGCA
951 CGGCGCGGGC AGTTCGCTTT ACGACCGCTT CAGCTGGTAT CCGAGGGCGG
1001 GCTTTCGAAGA AATCAAAACC GCCGAAAACC TGATCGGTAA AAAACCTGC
40 1051 CCAATTTTCG GCGGCGTGTG CGACAGCGAG CTGTTGCGCG AAGTGTGCGC
1101 ATTTTTCAAA AAACACGACA AGGGACTGTT TACTGGATG ACGCTGACCA
1151 GCCACGCCGA CTATCCCGAA TCCGACATTT TCAACCACAG GCTCAAATGC
1201 ACCGAATATG GCCTGCCCCG CGAAACCGAC CTCTGCCGCA ATTTACAGCT
1251 GCACACCCAA TTCTTCGACC AACTGGCGGA TTTGATCCAA CGCCCCGAAA
45 1301 TGAAGGCAC GGAAGTCATC ATCGTCGGCG ACCATCCGCC GCCCGTCGGC
1351 AACCTCAATG AAACCTTCCG CTACCTCAA CAGGGGCACG TCGCTGGCT
1401 GAACTTCAA ATCAAATAA
```

This corresponds to the amino acid sequence <SEQ ID 470; ORF48-1>:

```

50 1 MNIHTLLSKQ WTLPPFLPKR LLLSLILLA PNAVFWVLAL LTATARPIVN
51 LDYLPALLI ALPWRFKIA GVLAFWLAVL FDGLMMVIQL FPFMDLIGAI
101 NLVPFILTAP APYQIMTGLL LLYMLAMPFV LQKAAKTDF RHIAVCAAVV
151 AAAGYFTGHL SYDRGRMAN IFGANNFYA KSQAMLYTVS QNADFITAGL
201 VDPVFLPLGN QRAATHLNE PKSQKILFIV AESWGLPANP ELQNATFAKL
251 LAQKDRFSVW ESGSFPPFIGA TVEGEMRELC AYGGLRGFAL RRAPDEK FAR
55 301 CLPNRLKQEG YATFAMHGAG SSLYDRFSWY PRAGFQEIKT AENLIGKKT
```

351 AIFGGVCDSE LFGEVSAFFK KHDKGLFYWM TLTSHADYPE SDIFNHRCLK  
 401 TEYGLPAETD LCRNFSLHTQ FFDQLADLIQ RPEMKGTEVI IVGDHPPFPVG  
 451 NLNETFRYLK QGHVAVLNFK IK\*

Computer analysis of this amino acid sequence gave the following results:

5 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF48 shows 94.1% identity over a 119aa overlap with an ORF (ORF48a) from strain A of *N.*

*meningitidis*:

		10	20	30	40	50	60
10	orf48.pep	MNIHTLLSKQWTLPPFLPKRLLLSLLILLAPNAVFVWLALLTATARPVNLDPALPAALLI					
	orf48a	MNIHTLLSKQWTLPPFLPKRLLLSLLILLXPNVAVFWLALLTATARPVNLXYLPAALLI					
		10	20	30	40	50	60
15	orf48.pep	70	80	90	100	110	119
		ALPWRVFKIAGVLAFLAVLFDGLMMVLIQFPFMDLIGAINLVPFILTAPAPYQIMTGL					
	orf48a	ALPWRXVKIXGVLAFLAVLFDGLMMVLIQFPFMDLIGAINLVPFIXTAPALYQIMTGLL					
		70	80	90	100	110	120
20	orf48a	LLYMLAMPFVLQKAAAKTDFRHIACAAVVVAAGYFTGHLXSYDRGRMANIFGANNFYA					
		130	140	150	160	170	180

The complete length ORF48a nucleotide sequence <SEQ ID 471> is:

	1	ATGAATATTC	ACACCCTGCT	CTCCAAACAA	TGGACGCTGC	CGCCATTCCT
25	51	GCCGAAACGG	CTGCTGCTGT	CCCTGCTGAT	ACTGCTNNCC	CCCAATGCGG
	101	TGTTTGGGT	TTTGGCACTG	CTGACCGCCA	CCGCCGCCCC	GATTGTCAAT
	151	TTGGANTACC	TTCCGCGCGC	GCTGCTGATC	GCCCTGCCTT	GGCGTNTCGT
	201	CAAAATGNC	GGCGTATTGG	CGTNTTGGCT	GGCGGTTTTG	TTTGACGGGC
	251	TGATGATGGT	GATCCAATC	TTCCTTTTA	TGGATCTCAT	CGGCGCCATC
30	301	AACCTCGTCC	CCTTCACNT	GACCGCCCC	GCCCTTTATC	AGATAATGAC
	351	CGGGCTGTTA	CTGCTGTATA	TGCTGGCGAT	GCCGTTTGTG	TTGCAGAAAG
	401	CCGCCGCCAA	AACCGACTTC	CGACACATTG	CCGCCTGTGC	CGCCGTGTGT
	451	GTGGCAGCCG	GCTATTTTAC	CGGCCATTG	AGTTANTACG	ACCGGGGGCG
	501	AACGCGCAAT	ATCTTCGGCG	CAAACAATT	CTATTACGCC	AAAAGTCAGG
35	551	CGATGCTCTA	CACCGTCAGC	CAGAAATGCC	ACTTTATTAC	CGCCGCGCTG
	601	GTCGATCCCG	TCTTCTCC	CTTGGGCAAT	CAACAGCGTG	CCGCCACGCA
	651	TCTGAACGAG	CGGAAATCTC	AAAAAATCCT	CTTTATCGTC	GCCGAATCTT
	701	GGGGGCTGCC	GGCCAATCCC	GAACCTCAAA	ACGCCACTTT	TGCCAAACTG
	751	CTGGCGCAAA	AAGANCGTTT	TTCGGTTTGG	GAAAGCGGCA	GTTTTCCCTT
40	801	CATCGGCGCG	ACGATCGAAG	GCGAAATGCG	CGAACTGTGT	GCCTACGGCG
	851	GTTTGCGCGG	GTTTCGCACTG	CGCCGCGCGC	CCGACGAAAA	ATTTGCCCGC
	901	TGCCTCCCCA	ACCGTTTGAA	ACAAGAAGGT	TACGCCACCT	TTGCGATGCA
	951	CGGCGCGGGC	AGTTCGCTTT	ACGACCGCTT	CAGCTGGTAT	CCGAGGGCGG
	1001	GCTTTCAAGA	AATCAAAACC	GCCGAAAACC	TGATCGGTAA	AAAAACCTGC
	1051	GCCATTTTCG	GCGGCGTGTG	CGACAGCGAG	CTGTTCCGGC	AAGTGTCCGC
45	1101	ANTTTTCAAA	AAACACGACA	AGGGACTGTT	TTACTGGATG	ACGCTGACCA
	1151	GCCACGCCGA	CTATCCCGAA	TCNGACATTT	TCAACACAG	GCTCAATATG
	1201	ACCGAATATG	GCCTGCCCGC	CGAAACCGAC	NTCTGCCGCA	ATTTACGCCT
	1251	GCACACCCAA	TTCTTCGACC	AACTGGCGGA	TTTGATCCAA	CGCCCCGAAA
	1301	TGAAAGGCAC	GGAAGTCATC	ATCGTCGGCG	ACCATCCGCC	GCCCGTCGGC
50	1351	AACCTCAATG	AAACCTTCCG	CTACCTCAAA	CAGGGGCACG	TCGNCTGGCT
	1401	GAACCTCAAA	ATCAAAATAA			

This encodes a protein having amino acid sequence <SEQ ID 472>:

	1	MNIHTLLSKQ	WTLPPFLPKR	LLLSLLILLX	PNAVFWVLAL	LTATARPIVN
55	51	LXYLPAALLI	ALPWRXVKIX	GVLAXWLAVL	FDGLMMVIQL	FPPFMDLIGAI
	101	NLVPFXTAP	ALYQIMTGLL	LLYMLAMPFV	LQKAAAKTDF	RHIAACAAVV
	151	VAAGYFTGHL	SKYDRGRMAN	IFGANNFYA	KSQAMLYTVS	QNADFITAGL
	201	VDPVFLPLGN	QQRAATHLNE	PKSQKILFIV	AESWGLPANP	ELQNAFPAKL
	251	LAQKXRFVW	ESGSFPFIGA	TIEGEMRELC	AYGGLRGFAL	RRAPDEKFAR
	301	CLPNRLKQEG	YATFAMHGAG	SSLYDRFSWY	PRAGFQEIKT	AENLIGKKT
60	351	AIFGGVCDSE	LFGEVSAXFK	KHDKGLFYWM	TLTSHADYPE	SDIFNHRCLK

401 TEYGLPAETD XCRNFSLHTQ FFDQLADLIQ RPEMKGTEVI IVGDHPPPVG  
 451 NLNETFRYLK QGHVXWLNFK IK\*

ORF48a and ORF48-1 show 96.8% identity in 472 aa overlap:

5	orf48a.pep	MNIHTLLSKQWTLPPFLPKRLLLSLLILLXPNVFWVLALLTATARPIVNLXYLPAALLI	10	20	30	40	50	60
	orf48-1	MNIHTLLSKQWTLPPFLPKRLLLSLLILLAPNAVFWVLALLTATARPIVNLXYLPAALLI	10	20	30	40	50	60
10	orf48a.pep	ALPWRXVKIXGVLAVLFDGLMMVIQLFPFMDLIGAINLVFFIXTAPALYQIMTGLL	70	80	90	100	110	120
	orf48-1	ALPWRXVKIXGVLAVLFDGLMMVIQLFPFMDLIGAINLVFFIXTAPALYQIMTGLL	70	80	90	100	110	120
15	orf48a.pep	LLYMLAMPFVLQKAAAKTDFRHIACAAVVAAGYFTGHLSXYDRGRMANIFGANNFYA	130	140	150	160	170	180
	orf48-1	LLYMLAMPFVLQKAAAKTDFRHIACAAVVAAGYFTGHLSXYDRGRMANIFGANNFYA	130	140	150	160	170	180
20	orf48a.pep	KSQAMLYTVSQNADFITAGLVDPVFLPLGNQQRATHLNEPKSQKILFIVAESWGLPANP	190	200	210	220	230	240
	orf48-1	KSQAMLYTVSQNADFITAGLVDPVFLPLGNQQRATHLNEPKSQKILFIVAESWGLPANP	190	200	210	220	230	240
25	orf48a.pep	ELQNATFAKLLAQKXRFVSWESGSFFFIGATIEGEMRELCAYGGLRGFALRRAPDEKFA	250	260	270	280	290	300
	orf48-1	ELQNATFAKLLAQKXRFVSWESGSFFFIGATIEGEMRELCAYGGLRGFALRRAPDEKFA	250	260	270	280	290	300
30	orf48a.pep	CLPNRLKQEGYATFAMHGAGSSLYDRFSWYPRAGFQEIKAENLIGKKTCAIFGGVCDSE	310	320	330	340	350	360
	orf48-1	CLPNRLKQEGYATFAMHGAGSSLYDRFSWYPRAGFQEIKAENLIGKKTCAIFGGVCDSE	310	320	330	340	350	360
35	orf48a.pep	LFGEVSAXFKKHKDGLFYWMTLTSHADYPESDIFNHLKCTEYGLPAETDXCRNFSLHTQ	370	380	390	400	410	420
	orf48-1	LFGEVSAXFKKHKDGLFYWMTLTSHADYPESDIFNHLKCTEYGLPAETDXCRNFSLHTQ	370	380	390	400	410	420
40	orf48a.pep	FFDQLADLIQRPEMKGTEVIIVGDHPPPVGNLNETFRYLKQGHVXWLNFKIKX	430	440	450	460	470	
	orf48-1	FFDQLADLIQRPEMKGTEVIIVGDHPPPVGNLNETFRYLKQGHVXWLNFKIKX	430	440	450	460	470	

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF48 shows 97.5% identity over a 119aa overlap with a predicted ORF (ORF48ng) from *N.gonorrhoeae*:

55	orf48.pep	MNIHTLLSKQWTLPPFLPKRLLLSLLILLAPNAVFWVLALLTATARPIVNLXYLPAALLI	60
	orf48ng	MNIHTLLSKQWTLPPFLPKRLLLSLLILLAPNAVFWVLALLTATARPIVNLXYLPAALLI	60
60	orf48.pep	ALPWRXVKIXGVLAVLFDGLMMVIQLFPFMDLIGAINLVFFIXTAPALYQIMTGLL	119
	orf48ng	ALPWRXVKIXGVLAVLFDGLMMVIQLFPFMDLIGAINLVFFIXTAPALYQIMTGLL	120

The ORF48ng nucleotide sequence <SEQ ID 473> was predicted to encode a protein having amino acid sequence <SEQ ID 474>:

```

1  MNIHALLSEQ WTLPPFLPKR LLSLLILLA PNAVFWLAL LTATARPIVN
51 LDYLPALLI ALPWRFKIA GVLAFWPAVL FDGLMMVIQL FPFMDLIGAI
101 NLVPFILTAP APYQIMTGLL LLYMLAMPFV LQKAAVKTD F RHIAVCAAVV
151 AAARYFTGPF ELLRTGGRWQ YVQHRRLLS GSRASFRRRQ KADVLRLRGN
201 PYASMGNGG..

```

Further work identified the complete gonococcal DNA sequence <SEQ ID 475>:

```

1  ATGAATATTC ACGCCCTGCT CTCCGAACAA TGGACGCTGC CGCCATTCCT
10 51 GCCGAAACGG CTGCTGCTGT CCCTGCTGAT ACTGCTGGCC CCGAATGCGG
101 TGTTTTGGGT TTGGCACTG CTGACCGCCA CCGCCCGCCC GATTGTCAAT
151 TTGGACTACC TTCCCGCCG GCTGCTGATC GCCCTGCCTT GGCGTTTCGT
201 CAAAATTGCC GCGGTATTGG CGTTTTGGCC GCGGTTTTG TTTGACGGGC
15 251 TGATGATGGT GATCCAACTC TTCCCTTTTA TGGACCTCAT CGGCGGCATC
301 AACCTCGTCC CTTTCATCCT GACCGCCCCC GCCCCTTATC AGATAATGAC
351 CGGGCTGTTG CTGCTGTATA TGCTGGCGAT GCCGTTTG TGCAAAAAG
401 CCGCCGTCAA AACCGACTTC CGACACATTG CCGTCTGTGC CGCCGTTGTG
451 GCGGCAGCCG GCTATTTTAC CGGCCATTG AGTTACTACG ACCGGGGCGG
501 GATGGCCAAT ATCTTCGGCG CAAACAACCT CTATTACGCG aAAAGTCAGG
20 551 CGATGCTCTA CACCGTCAGC CAGAATGCCG ACTTTATTAC CGCCGgcctg
601 GTCGACCCCG TCTTCCTCCC CTGGGCAAT CAGCAGCGTG CCGCCACGCG
651 GCTGAGTGAG CCGAAATCTC AAAAAATCCT CTTTATCGTC GCCGAATCTT
701 GGGGGCTGCC GGGCAATCCC GAGCTTCAAA ACGCCACTTT TGCCAAACTG
751 CTGGCGCAAA AAGACCGTTT TTCGGTTTGG GAAAGCGGCA GTTTTCCCTT
25 801 CATCGGCGCG ACGGTCGAAG GCGAAATGCG CGAATTGTGC GCCTACGCGG
851 GTTTGCGCGG GTTCGCACTG CGCGCGCGCG CCGACGAAAA ATTTGCCCGC
901 TGCCTCCCA ACCGTTTGAA ACAAGAAGGT TACGCCACCT TTGCGATGCA
951 CGGCGCGGGT AGTTCGCTTT ACGACCGCTT CAGCTGGTAT CCGAGGGCGG
1001 GCTTTCAAAA AATCAAAACC GCCGAAAACC TGATCGGTAA AAAAACTGC
30 1051 GCCATTTTCG GCGGCGTGTG CGACAGCGAG CTGTTTCGGG AAGTGTGCGC
1101 ATTTTTTCAA AAACACGACA AGGGACTGTT TTAGTGATG ACGCTGACCA
1151 GCCACGCCGA CTATCCCGAA TCCGACATTT TCAACCACAG GCTCAAATGC
1201 ACCGAATACG GCCTGCCCGC CGAAACCGAC CTCTGCCGCA ATTTACGCTT
1251 GCACACCCAA TtcttegACC AACTGGCGGA TTTGATCCGA CGCCCCGAAA
35 1301 TGAAAGGCAC GGAAGTCATC ATCGTCGGCG ACCATCCGCC GCGCGTCGGC
1351 AACCTCAATG AAACCTTCG CTACCTCAAA CAGGGACACG TCGCTGGCT
1401 GCACTTCAAA ATCAATAA

```

This encodes a protein having amino acid sequence <SEQ ID 476; ORF48ng-1>:

```

1  MNIHALLSEQ WTLPPFLPKR LLSLLILLA PNAVFWLAL LTATARPIVN
40 51 LDYLPALLI ALPWRFKIA GVLAFWPAVL FDGLMMVIQL FPFMDLIGAI
101 NLVPFILTAP APYQIMTGLL LLYMLAMPFV LQKAAVKTD F RHIAVCAAVV
151 AAAGYFTGHL SYDRGRMAN IFGANFYFA KSQAMLYTVS QNADFITAGL
201 VDPVFLPLGN QQRAATRLSE PKSQKILFIV AESWGLPGNP ELQNAFPAKL
45 251 LAQKDRFSVW ESGSFPIGA TVEGEMRELC AYGLRGFAL RRAPDEKFA
301 CLPNRLKQEG YATFAMHGAG SLYDRFSWY PRAGFQIKT AENLIGKKT
351 AIFGGVCDSE LFGEVSFAFFK KHDKGLFYWM TLTSHADYPE SDIFNHLK
401 TEYGLPAETD LCRNFSLHTQ FFDQLADLIR RPEMKGTEVI IVGDHPPVPG
451 NLNETFRYLK QGHVAWLHFK IK*

```

ORG48ng-1 and ORF48-1 show 97.9% identity in 472 aa overlap:

```

50 orf48-1.pep      10      20      30      40      50      60
    MNIHTLLSKQWTLPPFLPKRLLLSLLILLAPNAVFWLALLTATARPIVNLDYLPALLI
    orf48ng-1      MNIHALLSEQWTLPPFLPKRLLLSLLILLAPNAVFWLALLTATARPIVNLDYLPALLI
55 orf48-1.pep      70      80      90      100     110     120
    ALPWRFKIAGVLAFWLAVLFDGLMMVIQLFPFMDLIGAINLVPFILTAPAPYQIMTGLL
    orf48ng-1      ALPWRFKIAGVLAFWPAVLFDGLMMVIQLFPFMDLIGAINLVPFILTAPAPYQIMTGLL
60 orf48-1.pep      70      80      90      100     110     120
    ALPWRFKIAGVLAFWLAVLFDGLMMVIQLFPFMDLIGAINLVPFILTAPAPYQIMTGLL
    orf48ng-1      ALPWRFKIAGVLAFWPAVLFDGLMMVIQLFPFMDLIGAINLVPFILTAPAPYQIMTGLL

```

-298-

		130	140	150	160	170	180
	orf48-1.pep	LLYMLAMPFVLQKAAAKTDFRHI	AVCAAVVAAAGYFTGHLSYYDRGR	MANIFGANNFY	YA		
	orf48ng-1	LLYMLAMPFVLQKAAVKTDFRHI	AVCAAVVAAAGYFTGHLSYYDRGR	MANIFGANNFY	YA		
5		130	140	150	160	170	180
	orf48-1.pep	KSQAMLYTVSQNADFITAGLVDPV	FLPLGNQQAATHLNEPKSQKILFIV	AESWGLPANP			
	orf48ng-1	KSQAMLYTVSQNADFITAGLVDPV	FLPLGNQQAATHLSEPKSQKILFIV	AESWGLPGNP			
10		190	200	210	220	230	240
	orf48-1.pep	ELQNATFAKLLAQKDRFSVWESGS	FPFIGATVEGEMRELCA	YGGLRGFALRRAPDEK	FAR		
	orf48ng-1	ELQNATFAKLLAQKDRFSVWESGS	FPFIGATVEGEMRELCA	YGGLRGFALRRAPDEK	FAR		
15		250	260	270	280	290	300
	orf48-1.pep	CLPNRLKQEGYATFAMHGAGSS	LYDRFSWYPRAGFQEI	KTAEENLIGKKTCAIF	GGVCDSE		
	orf48ng-1	CLPNRLKQEGYATFAMHGAGSS	LYDRFSWYPRAGFQEI	KTAEENLIGKKTCAIF	GGVCDSE		
20		310	320	330	340	350	360
	orf48-1.pep	LFGEVSAFFKKHDKGLFYWMT	LTSHADYPESDIFNHLRKCTEY	GLPAETDL	LCRNFSLHTQ		
	orf48ng-1	LFGEVSAFFKKHDKGLFYWMT	LTSHADYPESDIFNHLRKCTEY	GLPAETDL	LCRNFSLHTQ		
25		370	380	390	400	410	420
	orf48-1.pep	FFDQLADLIQRPEMKGTEV	IIVGDHPPVGNLNETFR	YLNKQGHV	AWLNEFKIKX		
	orf48ng-1	FFDQLADLIQRPEMKGTEV	IIVGDHPPVGNLNETFR	YLNKQGHV	AWLNEFKIKX		
30		430	440	450	460	470	
	orf48-1.pep	FFDQLADLIQRPEMKGTEV	IIVGDHPPVGNLNETFR	YLNKQGHV	AWLNEFKIKX		
	orf48ng-1	FFDQLADLIQRPEMKGTEV	IIVGDHPPVGNLNETFR	YLNKQGHV	AWLNEFKIKX		
35		430	440	450	460	470	

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and two putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### 40 Example 57

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 477>:

	1	..GTGAGCGGAC	GTTACCGCGC	TTTGGATCGC	GTTTCCAAAA	TCATCATCGT
	51	TACTTTGAGT	ATCGCCACGC	TTGCCGCCGC	CGGCATCGCT	ATGTCGCGCG
	101	GTATGCAGAT	GCACTCCGAT	TTTATCGAGC	CGACACCGTG	GACGCTTGCC
45	151	GGTTTGGGCT	TCCTGATCGC	GCTGATGGGC	TGGATGCCCG	CGCCGATTGA
	201	AATTTCCGCC	ATCAATTCTT	TGTGGGTAAC	CGAAAAACAA	CGCATCAATC
	251	CTTCCGAATA	CCGCGACGGG	ATTTTGAAT	TCAACGTCGG	TTATATCGCC
	301	AGTGCGGTTT	TGGCTTTGGT	TTTCCTTGCA	CTGGGCGC.G	TAGCGCGGAA
	351	CGGCAACGGC	GA.ACAGTGC	AGATGGCGGG	CGGCAAATAT	AACGGGCAAT
50	401	TGATCAATAT	GTACGCC..			

This corresponds to the amino acid sequence <SEQ ID 478; ORF53>:

	1	..VSGRYRALDR	VSKIIIVTLS	IATLAAAGIA	MSRGMQMOSD	FIEPTWTLA
	51	GLGFLIALMG	WMPAPIEISA	INSLWVTEKQ	RINPSEYRDG	IFEFNVGIIA
	101	SAVLALVFLA	LGXVAPNGNG	XTVQMAGGKY	NGQLINMYA..	

55 Further work revealed the complete nucleotide sequence <SEQ ID 479>:

1 ATGTCCGAAC AACATATTTTC GACTTGGAAA AGTAAATCA ACGCATTGGG



51 TCCGGGGATC ATGATGGCTT CGGCGGCGGT CGGCGGTTTC CACCTGATTG  
 101 CCTCGACGCA GGCGGGCGCG CTTTACGGCT GGCAGATCGC GCTCATCATC  
 151 ATCCTGACCA ACCTCTTCAA ATACCCGTTT TTCCGCTTCA GCGCGCATTA  
 201 CACGCTGGAC ACGGGCAAGA GCCTGATTGA AGGTTATGCC GAGAAAAGCC  
 251 GCGTTTATTT GTGGGTATTC CTGATTTTGT GCATCCTCTC CGCCACGATT  
 301 AACGCGGGCG CGGTCGCCAT TGTAACCGCC GCCATCGTCA AAATGGCGAT  
 351 TCCCTCGCTG ATGTTTGATG CCGGCACGGT TGCCGCCTTG ATTATGGCAT  
 401 CCTGCCTGAT TATTTTGGTG AGCGGACGTT ACCGCGCTTT GGATCGCGTT  
 451 TCCAAAATCA TCATCGTTAC TTGAGTATC GCCACGCTTG CCGCCGCGCG  
 501 CATCGCTATG TCGCGCGGTA TGCAGATGCA GTCCGATTTT ATCGAGCCGA  
 551 CACCGTGGAC GCTTGCCGGT TTGGGCTTCC TGATCGCGCT GATGGGCTGG  
 601 ATGCCCGCGC CGATTGAAAT TTCCGCCATC AATTCTTTGT GGGTAACCGA  
 651 AAAACAACGC ATCAATCCTT CCGAATACCG CGACGGGATT TTTGATTTCA  
 701 ACGTCGGTTA TATCGCCAGT GCGGTTTTGG CTTTGGTTTT CCTTGCACTG  
 751 GGCGCGTTTT TGCAATACCG CAACGGCGAA GCAGTGCAGA TGGCGGGCGG  
 801 CAAATATATC GGGCAATTGA TCAATATGTA CGCCGTACC ATCGGCGGCT  
 851 GGTGCGGCCC GCTGGTGGCG TTTATCGCGT TTGCCTGTAT GTACGGCACG  
 901 ACGATTACCG TCGTGGACGG CTATGCCCGT GCCATTGCCG AACCCGTGCG  
 951 CCTGCTGCGC GGAAAGACA AACCGGGCAA CGCCGAATTC TTTGCCTGGA  
 1001 ATATTTGGGT GGCGGGCAGC GGTTTGGCGG TGATTTTCTG GTTTGACGGC  
 1051 GTAATGGCGA ATCTGCTCAA ATTTGCGATG ATGCGCGCTT TGTGTCCGCG  
 1101 CCCTGTGTTT GCCTGGCTGA ATTACCGTTT GGTAAAGGT GATGAAAAAC  
 1151 ACAAACTCAC ATCAGGTATG AATGCCCTTG CATTGGCAGG CTTGATTTAT  
 1201 CTGACCGGTT TTACCGTTTT GTTCTTATTG AATTTGGCGG GAATGTTCAA  
 1251 ATGA

This corresponds to the amino acid sequence <SEQ ID 480; ORF53-1>:

1 MSEQHISTWK SKINALGPGI MMASAAVGS HLIASTQAGA LYGWQIALII  
 51 ILTNLFKYPF FRFSAHYTLT TGKSLIEGYA EKSRYVLWVF LILCILSATI  
 101 NAGAVAIUTA AIVKMAIPSL MFDAGTVAAL IMASCLIIIV SGRYRALDRV  
 151 SKIIIVTISI ATLAAAGIAM SRGMQMSDF IEPTPWLAG LGFLIALMGW  
 201 MPAPIEISAI NSLWVTEKQR INPSEYRDGI FDFNVGYIAS AVLALVFLAL  
 251 GAFVQYNGE AVQMAGGKYI GQLINMYAVT IGGWSRPLVA FIAFACMYGT  
 301 TITVVDGYAR AIAEPVRLLR GKDKTGNAEF FAWNIWVAGS GLAVIFWFDG  
 351 VMANLLKFAM IAAFVSAPVF AWLNYRLVKG DEKHKLTSKM NALALAGLIY  
 401 LTGFTVLFLL NLAGMFK\*

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF53 shows 93.5% identity over a 139aa overlap with an ORF (ORF53a) from strain A of *N. meningitidis*:

40 orf53.pep VSGRYRALDRVSKIIIVTLSIATLAAAGIA  
 orf53a AAIWKMAIPSLMFDAGTVAALIMASCLIIIVSGRYRALDRVSKIIIVTLSIATLAAAGIA  
 110 120 130 140 150 160  
 45 orf53.pep MSRGMQMSDFIEPTPWLAGLFLIALMGWMPAPIEISAINSLWVTEKQRINPSEYRDG  
 orf53a MSRGMQMSDFIEPTPWLAGLFLIALMGWMPAPIEISAINSLWVTEKQRINPSEYRDG  
 170 180 190 200 210 220  
 50 orf53.pep IFEFNVGYIASAVLALVFLALGXVAPNGNGXTVMAGGKYNGQLINMYA  
 orf53a IFDFNVGYIASAVLALVFLALGAFVQYNGEAVQMAGGKYIGQLINMYAVTIGGWSRPLV  
 230 240 250 260 270 280  
 orf53a AFIAFACMYGTTITVVDGYARAIAPVRLLRGKDKTGNAEFFAWNIWVAGSGLAVIFWFD  
 290 300 310 320 330 340

60 The complete length ORF53a nucleotide sequence <SEQ ID 481> is:

1 ATGTCCGAAC AACATATTTT GACTTGGAAG AGTAAATCA ACGCATTGGG  
 51 ACCGGGGATT ATGATGGCTT CGGCGGCGGT CGGCGGTTCG CACCTGATTG  
 101 CCTCGACGCA GCGGGGCGCG CTTTACGGCT GGCAGATCGC GCTCATCATC  
 151 ATCCTGACCA ACCTCTTCAA ATACCGTTT TCCGCTTCA GCGCGCATTA  
 201 CACGCTGGAC ACGGGCAAGA GCCTGATTGA AGGTTATGCC GAGAAAAGCC  
 251 GCGTTATTT GTGGGTATTC CTGATTTTGT GCATCCTCTC CGCCACGATT  
 301 AACGCGGGCG CGGTCCGCAT TGTAACCGCC GCCATCGTCA AAATGGCGAT  
 351 TCCCTCGCTG ATGTTTGATG CCGGCACGGT TGCCGCCTTG ATTATGGCAT  
 401 CCTGCTGAT TATTTTGGTG AGCGGACGTT ACCGCGCTT GGATCGCGTT  
 451 TCCAAAATCA TCATCGTTAC TTTGAGTATC GCCACGCTTG CCGCGCGCGG  
 501 CATCGCTATG TCGCGCGGTA TGCAGATGCA GTCCGATTT ATCGAGCCGA  
 551 CACCGTGGAC GCTTGCCGGT TTGGGCTTCC TGATCGCGCT GATGGGCTGG  
 601 ATGCCGCGCG CGATTGAAAT TTCCGCCATC AATTCTTTGT GGGTAACCGA  
 651 AAAACAACGC ATCAATCCTT CCGAATACCG CGACGGGATT TTTGATTCA  
 701 ACGTCGGTTA TATCGCCAGT GCGGTTTTGG CTTTGGTTTT CCTTGCACTG  
 751 GCGCGGTTTG TGCAATACGG CAACGGCGAA GCAGTGCAGA TGGCGGGCGG  
 801 CAAATATATC GGGCAATTGA TCAATATGTA CGCCGTTACC ATCGGCGGCT  
 851 GGTGCGGCCC GCTGGTGGCG TTTATCGCGT TTGCTGTAT GTACGGCACG  
 901 ACGATTACCG TTGTGGACGG CTATGCCCGT GCCATTGCCG AACCCGTGCG  
 951 CCTGCTGCGC GGAAAAGACA AAACGGGCAA CGCGGAATTC TTTGCTGGA  
 1001 ATATTGGGT GCGGGGCGAG GGTTTGGCGG TGATTTCTG GTTTGACGGC  
 1051 GTAATGGCGA ATCTGCTCAA ATTTGCGATG ATTGCGGCT TGTGTCCGC  
 1101 CCCTGTGTTT GCCTGGCTGA ATTACCGTTT GGTCAAAGGT GATGAAAAAC  
 1151 ACAAACCTAC ATCAGGTATG AATGCCCTTG CATTTGGCAGG CTTGATTTAT  
 1201 CTGACCGGTT TTACCGTTTT GTTCTTATTG AATTGGCGG GAATGTTCAA  
 1251 ATGA

This encodes a protein having amino acid sequence <SEQ ID 482>:

1 MSEQHISTWK SKINALGPGI MMASAAVGGSHLIASTQAGA LYGWQIALII  
 51 ILTNLFKYPF FRFSAHYTLD TGKSLIEGYA EKSRYVLWVF LILCILSATI  
 101 NAGAVAIUTA AIVKMAIPSL MFDAGTVAAL IMASCLIIIV SGRYRALDRV  
 151 SKIIIVTSLI ATLAAAGIAM SRGMQMOSDF IEPTPWTLAG LGFLIALMGW  
 201 MPAPIEISAI NSLWVTEKQR INPSEYRDGI FDFNVGYIAS AVLALVFLAL  
 251 GAFVQYNGE AVQMAGGKYI GQLINMYAVT IGGWSRPLVA FIAFACMYGT  
 301 TITVVDGYAR ATAEPVRLLR GKDKTGNAEF FAWNIVWAGS GLAVIEWFDG  
 351 VMANLLKFAM IAAFVSAFVF AWLNYRLVKG DEKHKLTSGM NALALAGLIY  
 401 LTGFTVLFLL NLAGMFK\*

ORF 53a shows 100.0% identity in 417 aa overlap with ORF53-1:

		10	20	30	40	50	60
40	orf53a.pep	MSEQHISTWKS	SKINALGPGI	MMASAAVGGSH	LIASTQAGALY	GWQIALII	ILTNLFKYPF
	orf53-1	MSEQHISTWKS	SKINALGPGI	MMASAAVGGSH	LIASTQAGALY	GWQIALII	ILTNLFKYPF
		10	20	30	40	50	60
45	orf53a.pep	FRFSAHYTLD	TGKSLIEGYA	EKSRYVLWVF	LILCILSATI	NAGAVAIUTA	AIVKMAIPSL
	orf53-1	FRFSAHYTLD	TGKSLIEGYA	EKSRYVLWVF	LILCILSATI	NAGAVAIUTA	AIVKMAIPSL
		70	80	90	100	110	120
50	orf53a.pep	MFDAGTVAAL	IMASCLIIIV	SGRYRALDRV	SKIIIVTSLI	ATLAAAGIAM	SRGMQMOSDF
	orf53-1	MFDAGTVAAL	IMASCLIIIV	SGRYRALDRV	SKIIIVTSLI	ATLAAAGIAM	SRGMQMOSDF
		130	140	150	160	170	180
55	orf53a.pep	IEPTPWTLAG	LGFLIALMGW	MPAPIEISAI	NSLWVTEKQR	INPSEYRDGI	FDFNVGYIAS
	orf53-1	IEPTPWTLAG	LGFLIALMGW	MPAPIEISAI	NSLWVTEKQR	INPSEYRDGI	FDFNVGYIAS
		190	200	210	220	230	240
60	orf53a.pep	AVLALVFLAL	GAFVQYNGE	AVQMAGGKYI	GQLINMYAVT	IGGWSRPLVA	FIAFACMYGT
	orf53-1	AVLALVFLAL	GAFVQYNGE	AVQMAGGKYI	GQLINMYAVT	IGGWSRPLVA	FIAFACMYGT
		250	260	270	280	290	300
65	orf53a.pep	AVLALVFLAL	GAFVQYNGE	AVQMAGGKYI	GQLINMYAVT	IGGWSRPLVA	FIAFACMYGT
	orf53-1	AVLALVFLAL	GAFVQYNGE	AVQMAGGKYI	GQLINMYAVT	IGGWSRPLVA	FIAFACMYGT
		250	260	270	280	290	300

		310	320	330	340	350	360
	orf53a.pep	TITVVDGYARAIAEPVRLLRGKDKTGNAEFFAWNIVVAGSGLAVIFWFDGVMANLLKFAM					
5	orf53-1	TITVVDGYARAIAEPVRLLRGKDKTGNAEFFAWNIVVAGSGLAVIFWFDGVMANLLKFAM					
		310	320	330	340	350	360
	orf53a.pep	IAAFVSAPVFAWLNYRLVKGDEKHKLTSGMNALALAGLIYLTGFTVLFLLNLAMGFKX					
10	orf53-1	IAAFVSAPVFAWLNYRLVKGDEKHKLTSGMNALALAGLIYLTGFTVLFLLNLAMGFKX					
		370	380	390	400	410	

### Homology with a predicted ORF from *N.gonorrhoeae*

- 15 ORF53 shows 92.1% identity over a 139aa overlap with a predicted ORF (ORF53ng) from *N. gonorrhoeae*:

	orf53.pep	VSGRYRALDRVSKIIIVTLSIATLAAAGIA	30
	orf53ng	AAIVKMAIPSLMFDAGTVAALIMASCLIIIVSRYRALDRVSKIIIVTLSIATLAAAGIA	91
20	orf53.pep	MSRGMQMOSDFIEPTPWTLAGLGLIALMGWMPAPIEISAINSLWVTEKQRINPSEYRDG	90
	orf53ng	MSRGMQMOPDFIEPTPWTLAGLGLIALMGWMPAPIEISAINSLWVTEKQRINPSEYRDG	151
25	orf53.pep	IFEFNVGYIASAVLALVFLALGXVAPNGNGXTVMAGGKYNGQLINMYA	139
	orf53ng	IFDFNVGYIASAVLALVFLALGAFVQYNGEAVQMGGGKYIGQLINMYAVTIGGSRPLV	211

An ORF53ng nucleotide sequence <SEQ ID 483> was predicted to encode a protein having amino acid sequence <SEQ ID 484>:

	1	MPKKSCVYLW VFLILCIASA TINAGAVAIV TAAIVKMAIP SLMFDAGTVA
	51	ALIMASCLII LVSGRYRALD RVSKIIIVTL SIATLAAAGI AMSRGMQMOP
	101	DFIEPTPWTLAGLGLIALMGWMPAPIEIS AINSLWVTEK QRINPSEYRD
	151	GIFDFNVGYI ASAVLALVFL ALGAFVQYGN GEAVQMGGGK YIGQLINMYA
35	201	VTIGGSRPL VAFIAFACMY GAASTVVDGY ARAIAEPVRL LRKDKTARP
	251	IVLLEKLGGH HRFGRDFLV*

Further analysis revealed further partial DNA gonococcal sequence <SEQ ID 485>:

	1	..aagaAAAGCT GCGTTTATTT GTGGGTTTTT TTGATTTTGT GTATCGCCTC
	51	CGCCACGATT AACGCGGGCG CGGTCGCCAT TGTAACCGCC GCCATCGTCA
40	101	AAATGGCGAT TCCCTCGCTG ATGTTTGATG CCGGCACGGT TGCCGCTTG
	151	ATTATGGCAT CCTGCCTGAT TATTTTGGTG AGCGGACGTT ACCGCGCTTT
	201	GGATCGTGTT TCCAAAATCA TCATTGTTAC TTTGAGCATC GCCACGCTTG
	251	CCGCCGCGG CATCGCTATG TCGCGCGGTA TGCAGATGCA GCCCGATTTT
45	301	ATCGAGCCGA CACCGTGGAC GCTTGCCGGT TTGGGCTTCC TGATCGCGCT
	351	GATGGGCTGG ATGCCGCGCG CGATCGAAAT TTCCGCCATC AATTCTTTGT
	401	GGGTAACCGA AAAACAACGC ATCAATCCTT CTGAATACCG CGACGGGATT
	451	TTGATTTTCA ACGTCGGTTA TATCGCagT GCGGTTTTGG CTTTGGTTTT
	501	CCTTGCACTG GCGCGGTTTG TGCAATACGG CAACGCGGAA GCAGTGCGA
50	551	TGGCGGGCGG CAAATATATC GGGCAATTGA TTAATATGTA TGCCGTAACC
	601	ATCGGCGGCT GGTCTCGTCC GCTGGTGGCG TTTATCGCGT TTGCCTGTAT
	651	GTACGGCAGC ACATTACCG TTGTGGACGG TTATGCGCGT GCCATTGCCG
	701	AACCCGTGCG CCTGCTGCGC GGCAGGGATA AAACCGGCAA CGCCGAGTTG
	751	TTtgccTGGA ATATTTGGGT GCGGGGAGC GGTTTGGCGG TGATTTTCTG
55	801	GTTTGACggc gcaaTGGCGg AACTgctCAA ATTTGCGATG AttgccgcCT
	851	TTGTGTCCGC CCTGTGTTC GCCTGGCTCA ACTACCGCTT CGTCAAAGGG
	901	GACAAACGCC ACAGGCTTAC CGCCGGTATG AACGCCCTTG CCATTGTCCG
	951	CCTGCTCTAC CTGCGGGGT TTGCCGTTTT GTTCCTGTG AACCTTACCG
	1001	GACTTTTGGC ATAG

This corresponds to the amino acid sequence <SEQ ID 486; ORF53ng-1>:

1 ..KKSCVYLWVF LILCIASATI NAGAVAIVTA AIVKMAIPSL MFDAGTVAAL  
 51 IMASCLIIILV SGRYRALDRV SKIIIVTSLI ATLAAAGIAM SRGMQMOPDF  
 101 IEPTPWTLAG LGFLIALMGW MPAPIEISAI NSLWVTEKQR INPSEYRDGI  
 151 FDFNVGYIAS AVLALVFLAL GAFVQYNGE AVQMAGGKYI GOLINMYAVT  
 5 IGGWSRPLVA FIAFACMYGT TITVVDGYAR AIAEPVRLLR GRDKTGNAEL  
 251 FAWNIWVAGS GLAVIFWFDG AMAELLKFAM IAAFSAPVF AWLNYRLVKG  
 301 DKRHRLTAGM NALAIVGLLY LAGFAVLFLN NLTGLLA\*

ORF53ng-1 and ORF53-1 show 94.0% identity in 336 aa overlap:

10	orf53-1.pep	ILTNLFKYPPFRFSAHYTLDTGKSLIEGYAEKSRVYLWVFLILCILSATINAGAVAIVTA	60	70	80	90	100	110
	orf53ng-1	KKSCVYLWVFLILCIASATINAGAVAIVTA				10	20	30
15	orf53-1.pep	AIVKMAIPSLMFDAGTVAALIMASCLIIILVSGRYRALDRVSKIIIVTSLIATLAAAGIAM	120	130	140	150	160	170
	orf53ng-1	AIVKMAIPSLMFDAGTVAALIMASCLIIILVSGRYRALDRVSKIIIVTSLIATLAAAGIAM						
20	orf53-1.pep	SRGMQMOSDFIEPTPWTLAGLGFLIALMGWMPAPIEISAINSLWVTEKQRINPSEYRDGI	180	190	200	210	220	230
	orf53ng-1	SRGMQMOPDFIEPTPWTLAGLGFLIALMGWMPAPIEISAINSLWVTEKQRINPSEYRDGI						
25	orf53-1.pep	FDFNVGYIASAVLALVFLALGAFVQYNGEAVQMAGGKYIGQLINMYAVTIGGWSRPLVA	240	250	260	270	280	290
	orf53ng-1	FDFNVGYIASAVLALVFLALGAFVQYNGEAVQMAGGKYIGQLINMYAVTIGGWSRPLVA						
30	orf53-1.pep	FIAFACMYGTTITVVDGYARAIAEPVRLLRGDKGTGNAEFFAWNIWVAGSGLAVIFWFDG	300	310	320	330	340	350
	orf53ng-1	FIAFACMYGTTITVVDGYARAIAEPVRLLRGRDKGTGNAELFAWNIWVAGSGLAVIFWFDG						
35	orf53-1.pep	VMANLLKFAMIAAFVSAPVFAWLNYRLVKGDEKHKLTSGMNALALAGLIYLTGFTVLFLN	360	370	380	390	400	410
	orf53ng-1	AMAELLKFAMIAAFVSAPVFAWLNYRLVKGDKRHRLTAGMNALAIVGLLYLAGFAVLFLN						
40	orf53-1.pep	NLAGMFKX	280	290	300	310	320	330
	orf53ng-1	NLTGLLAX						

50 Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 58

55 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 487>:

1 ..TTGCGGGA AAA CGGCATATGT TTTGGATAGT TTTGATCGTT ATTTTGTGT  
 51 TGCGCTTGCC GGCTTGTTTT TTGTCCGCGC ACAATCCGAA CGCGAGTGGA  
 101 TGCGCGAGGT TTCTGCGTGG CAGGAAAAGA AAGGGGAAAA ACAGGCGGAG

151 CTGCTTGAAG TCAAAGACGG TATGCCCGAT TTTCCCGAAC TTGCCCTGAT  
201 GCTTTTCCAC GCCGTCAAAA CGGCAGTGTA TTGGCTGTTT GTCGGTGTGG  
251 TCCGTTTCTG CCGAAACTAT CTGGCGCACG AATCCGAACC GGACAGGCCG  
301 GTTCCGCCT..

5 This corresponds to the amino acid sequence <SEQ ID 488; ORF58>:

1 ..LRETAYVLDS FDRYFVVALA GLFFVRAQSE REWMREVSAA QEKKGEKQAE  
101 LPEIKDGMFD FPELALMLFH AVKTAVYWLF VGVVRFRCRNY LAHESEPDRP  
VPP..

Further work revealed the complete nucleotide sequence <SEQ ID 489>:

10 1 ATGTTTTGGA TAGTTTTGAT CGTTATTTTG TTGCTTGCGC TTGCCCGCTT  
51 GTTTTTTGTG CGCGCACAAAT CCGAACGCGA GTGGATGCGC GAGGTTTCTG  
101 CGTGGCAGGA AAAGAAAGGG GAAAAACAGG CGGAGCTGCC TGAAATCAAA  
151 GACGGTATGC CCGATTTTCC CGAACTTGCC CTGATGCTTT TCCATGCCGT  
201 CAAACGGCA GTGTATTGGC TGTTTGTCGG TGTCGTCCGT TTCTGCCGAA  
15 251 ACTATCTGGC GCACGAATCC GAACCGGACA GGCCCGTTCC GCCTGCTTCT  
301 GCAAACCGTG CGGATGTTCC GACGCGATCC GACGGATATT CAGACAGTGG  
351 AAACGGGACG GAAGAAGCGG AAACGGAAGA AGCAGAAGCT GCGGAGGAAG  
401 AGGCTGCCGA TACGGAAGAC ATTGCAACTG CCGTAATCGA CAACCCGCCG  
451 ATCCCATTCG ACCGGAGTAT TGCTGAAGGG TTGATGCCGT CTGAAAGCGA  
20 501 AATTTCCGCC GTCCGTCCGG TTTTAAAGA AATCACTTTG GAAGAAGCAA  
551 CGCGTGCTTT AAACAGCGCG GCTTTAAGGG AAACGAAAAA ACGCTATATC  
601 GATGCATTGG AGAAAAACGA AACAGCGGTC CCCAAAGTCC GCGTGTCCGA  
651 TACCCCGATG GAAGGGCTGC AGATTATCGG TTTGGACGAC CCTGTGCTTC  
701 AACGCACGTA TTCCCATATG TTCGATGCGG ACAAAGAAGC GTTTTCCGAG  
25 751 TCTGCGGATT ACGGATTTGA GCCGATTTT GAGAAGCAGC ATCCGTCTGC  
801 CTTTCTGCA GTCAAAGCCG AAAATGCACG GAATGCGCCG TTCCACCGTC  
851 ATGCAGGGCA GGGGAAAGGG CAGGCGGAGG CAAAATCCCC GGATGTTTCC  
901 CAAGGGCAGT CCGTTTCAGA CGGCACGGCC GTCCCGCATG CCCGCCGCCG  
951 CGTTTCCGTC AATTGAAAG AACCGAACAA GGCAACGGTT TCTGCGGAGG  
30 1001 CGCGAATTTT TCGCCTGATT CCGGAAAGTC AGACGGTTGT CGGGAACCGG  
1051 GATGTCGAAA TGCGTCTGA AACCGAAAT GTTTTACGG AAACCGTTTC  
1101 GTCTGTGGGA TACGGCGGTC CGTTTATGA TGAACCTGCC GATATCCATA  
1151 TTGAAGAAC TGCCGCGCCC GATGCTTGGG TGTCGAACC ACCCGAAGTG  
35 1201 CCGAAAGTTC CCATGACCGC AATCGATATT CAGCCGCGC CTCCCGTATC  
1251 GGAAATCTAC AACCGTACCT ATGAACCGCC GTCAAGATTG GAGCAGGTGC  
1301 AACGCAGCCG CATTGCCGAG ACCGACCATC TTGCCGATGA TGTTTGAAT  
1351 GGAGGTTGGC AGGAGGAAAC CGCCGCTATT GCGGATGACG GCAGTGAAGG  
1401 TGCGGCAGAG CGGTCAAGCG GGCAATATCT GTCGGAAACC GAAGCGTTTCG  
1451 GGCATGACAG TCAGGCGGTT TGTCGTTTG AAAATGTGCC GTCTGAACGC  
40 1501 CCGTCTGTC GGGTATCGGA TACGGAAGCG GATGAAGGGG CGTTCCTATC  
1551 TGAAGAAACC GGTGCGGTAT CCGAACACCT GCCGACAACC GACCTGCTTC  
1601 TGCTTCCGCT GTTCAATCCC GAGGCGACGC AAACCGAAGA AGAACTGTTG  
1651 GAAACAGCA TCACCATCGA AGAAAAATG GCGGAGTTCA AAGTCAAGGT  
1701 CAAGGTTGTC GATTCTTATT CCGGCCCGT AATTACGCGT TATGAAATCG  
45 1751 AACCCGATGT CGGCGTGC GCATTTCCG TTTCTGAATCT GGAAAAAGAT  
1801 TTGGCGCGTT CGCTCGGCGT GGCTTCCATC CGCGTTGTCG AAACCATCCC  
1851 CGGCAAAACC TGCAATGGGT TGGAACTTCC GAACCCGAAA CGCCAAATGA  
1901 TACGCCTGAG CGAAATCTTC AATTCGCCG AGTTTCCGA ATCCAAATCC  
1951 AAGCTGACGC TCGCGCTCGG TCAGGACATC ACCGGACAGC CCGTCTGTAAC  
50 2001 CGACTTGGGA AAAGCACC GCATTTGTTGTT TGCCGGCAGC ACCGGTTCGG  
2051 GCAATTCGGT GGGTGTCAAC GCGATGATTC TGCTATGCT TTTCAAAGCC  
2101 GCGCCGGAAG ACGTGCGTAT GATTATGATC GATCCGAAA TGCTGGAATT  
2151 GAGCATTTAC GAAGGCATCC CGCACCTGCT CGCCCTGTC GTTACCGATA  
2201 TGAAGCTGGC GGCAACGCG CTGAACCTGT GTGTTAACGA AATGGAAGAA  
55 2251 CGCTACCGCC TGATGAGCTT TATGGCGTG CGTAATCTTG CGGGCTTCAA  
2301 TCAGGAAATC GCCGAAGCCG CAGCAAGGGG AGAAAAATC GGCAATCCGT  
2351 TCAGCTCAC GCGGACGAT CCGCAACCTT TGGAAGAACT GCGGTTTATC  
2401 GTGGTCGTGG TCGATGAGTT TGCCGACCTG ATGATGACGG CAGGCAAGAA  
2451 AATCGAAGAA CTGATTGCC GCCTCGCCA AAAAGCCCCG GCGGAGGCA  
60 2501 TCCATTTGAT TCTTGCCACA CAACGCCCA GCGTCGATGT CATCACGGGT  
2551 CTGATTAAGG CGAATATCCC GACGCGTATC GCGTTCCAAG TGTCAGCAA  
2601 AATCGACAGC CGCACGATTC TCGACCAAT GGGCGCGGAA AACCTGCTCG  
2651 GTCAGGGCGA TATGCTGTTT CTGCTGCCG GTACTGCCTA TCCGAGCCG  
2701 GTTACGGCG CGTTTGCTTC GGATGAAGAG GTGCACCGCG TGGTCGAATA  
65 2751 TTTGAAACAG TTTGCGCAAC CGGACTATGT TGACGATATT TTGAGCGGGC  
2801 GCGGACGCGA AGAGCTGCC GGCATCGGCG GCAGCGGCGA CGACGAACCC

2851 GATCCGATGT ACGACGAGGC CGTATCCGTT GTCCTGAAAA CGCGCAAAGC  
 2901 CAGCATTTTCG GCGGTACAGC GCGCCTTGCG TATCGGCTAC AACCAGCGCCG  
 2951 CGCGTCTGAT TGACCAGATG GAGGCGGAAG GCATTGTGTC CGCACCAGGAA  
 3001 CACAACGGCA ACCGTACGAT TCTCGTCCCC TTGGACAATG CTGGA

5 This corresponds to the amino acid sequence <SEQ ID 490; ORF58-1>:

1 MFWIVLVIL LLALAGLFFV RAQSEREWMR EVSAWQEKKG EKQAELEPEIK  
 51 DGMPDFPELA LMLFHAVKTA VYWLFGVVVR FCRNYLAHES EPDRPVPPAS  
 101 ANRADVPTAS DGYSDSGNGT EEAETEEAEA AEEEAADTED IATAVIDNRR  
 151 IPFDRSIAEG LMPSESEISP VRPVFKEITL EEATRALNSA ALRETKKRYI  
 201 DAFEKNETAV PKVRVSDTPM EGLQIIGLDD PVLQRTYSHM FDADKEAFSE  
 251 SADYGFEPYF EKQHPSAFSA VKAENARNAP FHRHAGQKGQ QAEAKSPDVS  
 301 QGQSVSDGTA VRDARRRVSV NLKEPNKATV SAEARISRLI PESQTVVGKR  
 351 DVEMPSETEN VFTETVSSVG YGGPVYDETA DIHIEEPAAP DAWVVEPEV  
 401 PKVPMTAIDI QPPPPVSEIY NRTYEPPSGF EQVQRSRIAE TDHLADDVLN  
 451 GGWQEETAAL ADDGSEGAAE RSSGQYLSET EAFGHDSQAV CPFENVPSER  
 501 PSCRVSDETA DEGAFPSSET GAVSEHLPTT DILLPPLFNP EATQTEEELL  
 551 ENSITIEEKL AEFKVKVKV DSYSGPVITR YEIEPDVGVV GNSVLNLEKD  
 601 LARSLGVASI RVVETIPGKT CMGLELPNPK RQMIRLSEIF NSPEFAESKS  
 651 KLTALALGQDI TGQPVVTDLG KAPHLVAGT TSGGKSVGVN AMILSMLFKA  
 701 APEDVRMIMI DPKMLELSIY EGIPHLLAPV VTDMLKLAANA LNWCVNEMEK  
 751 RYRLMSFMGV RNLAGFNQKI AEAARGEKI GNPFSLTPDD PEPEKLPFI  
 801 VVVVDEFADL MMTAGKKIEE LIARLAQKAR AAGIHLILAT QRPSVDVITG  
 851 LIKANIPTRI AFQVSSKIDS RTILDQMGAE NLLGQGDMLF LLPGTAYPQR  
 901 VHGAFAASDEE VHRVVEYLKQ FGEPDYVDDI LSGGGSEELP GIGRSGDDET  
 951 DPMYDEAVSV VLKTRKASIS GVQRALRIGY NRAARLIDQM EAEGIVSAPE  
 1001 HNGNRTILVP LDNA\*

Computer analysis of this amino acid sequence predicts the indicated transmembrane region, and also gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

30 ORF58 shows 96.6% identity over a 89aa overlap with an ORF (ORF58a) from strain A of *N. meningitidis*:

	10	20	30	40	50	60
orf58.pep	LRETAYVLDSFDRYFVVALAGLFFVRAQSEREWMR	EVSAWQEKKG	EKQAELEPEIK	DGMPDFPELA		
	:::					
orf58a	MFWIVLVILLLALAGLFFVRAQSEREWMR	EVSAWQEKKG	EKQAELEPEIK	DGMPDFPELA		
	10	20	30	40	50	
	70	80	90	100		
orf58.pep	FPPELALMLFHAVKTAVYWLFGVVVRCRNYLAHES	EPDRPVPP				
orf58a	FPPELALMLFHAVKTAVYWLFGVVVRCRNYLAHES	EPDRPVPPAS	ANRADVPTAS	DGYS		
	60	70	80	90	100	110

The complete length ORF58a nucleotide sequence <SEQ ID 491> is:

1 ATGTTTTGGA TAGTTTTGAT CGTATTTTGG TTGCTTGCGC TTGCCGGCTT  
 45 51 GTTTTTTGTC CGGCACAAAT CCGAACGCGA GTGGATGCGC GAGGTTTCTG  
 101 CGTGGCAGGA AAAGAAAGGG GAAAAACAGG CGGAGCTGCC TGAATCAAA  
 151 GACGGTATGC CCGATTTTCC CGAACTTGCC CTGATGCTTT TCCATGCCGT  
 201 CAAAACGGCA GTGTATTGGC TGTTTGTCGG TGTCGTCCGT TTCTGCCGAA  
 251 ACTATCTGCG GCACGAATCC GAACCGGACA GGCCCGTTCC GCCTGCTTCT  
 50 301 GCAAATCGTG CGGATGTTCC GACCGCATCC GACGGATATT CAGACAGTGG  
 351 AAACGGGACG GAAGAAGCGG AAACGGGAAGA AGCAGAAGCT GCGGAGGAAG  
 401 AGGCTGCCGA TACGGAAGAC ATTGCAACTG CCGTAATCGA CAACCGCCGC  
 451 ATCCCATTCG ACCGGAGTAT TGCTGAAGGG TTGATGCCGT CTGAAAGCGA  
 501 AATTTCCGCC GTCCGTCCGG TTTTAAAGGA AATCACTTTG GAAGAAGCAA  
 55 551 CGCGTGCTTT AAACAGCGCG GCTTTAAGGG AAACGAAAAA ACGTATATC  
 601 GATGCATTG AGAAAAACGA AACAGCGGTC CCCAAAGTCC GCGTGTCCGA  
 651 TACCCCGATG GAAGGGCTGC AGATTATCGG TTTGGACGAC CCTGTGCTTC  
 701 AACGCACGTA TTCCCGTATG TTCGATGCGG ACAAAGAAGC GTTTTCCGAG

5  
10  
15  
20  
25  
30  
35  
40  
45

```

751 TCTGCGGATT ACGGATTTGA GCCGTATTTT GAGAAGCAGC ATCCGTCTGC
801 CTTTTCTGCA GTCAAAGCCG AAAATGCACG GAATGCGCCG TTCCGCGGTC
851 ATGCAGGGCA GGNAAAGGG CAGGCGGAGG CNAATCCCC GGATGTTTCC
901 CAAGGGCAGT CCGTTTCAGA CGGCACAGCC GTCCGCGATG CCNGCCCGCG
951 CGTTTCCGTC AATTTGAAAG AACCGAACAA GGCAACGGTT TCTGCGGAGG
1001 CGCGGATTTC GCGCCTGATT CCGGAAAGTC GGACGGTTGT CGGGAAACGG
1051 GATGTCGAAA TGCCGTCTGA AACCGAAAAT GTTTTCACGG AAANTGTTTC
1101 GTCTGTGGGA TACGGCGNTC CGGTTTATGA TGAAACTGCC GATATCCATA
1151 TTGAAGAACC TGCCGCGCCC GATGCTTGGG TGGTCAAGCC ACCCGAAGTG
1201 CCGAAAGTTC CCATGCCCGC AATNGATATT CCGCCGCGCG CTCCCCGTATC
1251 GGAATCTTAC AACCGTACCT ATGAACCGCC GGCAGGATTG GAGCAGGTGC
1301 AACGCAGCCG CATTGCCGAA ACCGATCATC TTGCCGATGA TGTTTTGAAT
1351 GGAGGTTGGC AGGAGGAAAC CGCCGCTATT GCGAATGACG CGAGTGAGGG
1401 TGTGGCAGAG CGGTCAAGCG GGCAATATTT GTCGGAACCC GAAGCGTTTCG
1451 GGCATGACAG TCAGGCGGTT TGTCCGTTTG AAAATGTGCC GTCTGAACGC
1501 CCGTCCCGCC GGGCATNGGA TACGGAAGCG GATGAAGGGG CGTTCCAATC
1551 TGAAGAAACC GGTGCGGTAT CCGAACACCT GCCGACAAAC GACCTGCTTC
1601 TGCCGCGGCT GTTCAATCCC GGGGCGACGC AAACCGAAGA AGANCTGTTG
1651 GANAACAGCA TCACCATCGA AGAAAAATNG GCGGAGTTCA AAGTCAAGGT
1701 CAAGGTTGTC GATTCTTATT CCGGCCCCGT GATTACGCGT TATGAAATCG
1751 AACCCGATGT CCGCGTGCGC GGCAATTCGG TTCTAAATCT GGAAAAAGAN
1801 TTGGCGCGTT CGCTCGGCGT GGCTTCCATC CGCGTTGTCT AAACCATCTT
1851 CGGCAAAACC TGTATGGGTT TGGAACCTCC GAACCCGAAA CGCCAAATGA
1901 TACGCTGAGC CGAAATCTTC AATTCGCCCG AGTTTGCCGA ATCCAAATCC
1951 AAGCTGACGC TCGCGCTCGG TCAGGACATC ACCGGACAGC CCGTCGTAAC
2001 CGACTTGGGC AAAGCACCGC ATTTGTTGGT TGCCGGCAGC ACCGGTTCGG
2051 GCAAATCGGT GGGTGTCAAC GCGATGATTC TGTCTATGCT TTTCAAAGCC
2101 GCGCCGGAAG ACGTGCGTAT GATTATGATC GATCCGAAAA TGCTGGAATT
2151 GAGCATTTAC GAAGGCATCC CGCACCTGCT CGCCCCGTGC GTTACCGATA
2201 TGAAGCTGGC GGCAAACCGC CTGAACCTGT GTGTTAACGA AATGGAAGAA
2251 CGCTACCGCC TGATGAGCTT TATGGGCGTG CGCAATCTTG CGGGTNTCAA
2301 TCAAAAAATC GCCGAAGCCG CAGCAAGGGG GGAGAAAATC GGCAACCCGT
2351 TCAGCCTCAC GCGGACAAT CCCGAACCTT TGGANAATT GCGTTTATC
2401 GTGTCGTGG TTGATGAGTT TGCCGACCTG ATGATGACGG CAGGCAAGAA
2451 AATCGAAGAA CTGATTGCCC GCCTCGCCCA AAAAGCCCGC GCGGCAGGCA
2501 TCCATCTTAT CCTTGCCACA CAACGCCCCA GTGTCGATGT CATCAGGGT
2551 CTGATTAAGG CGAACATCCC GACGCGTATC GCGTTCCAAG TGTCCAGCAA
2601 AATCGACAGC CGCAGGATTC TTGACCAAAT GGGTGCGGAA AACCTGCTCG
2651 GGCAGGGCGA TATGCTGTTT CTGCCGCCGG GTACGGCCTA TCCGCAGCGC
2701 GTTCACGGCG CGTTTGCCCT GGATGAAGAG GTGCAACCGG TGGTCAAGTA
2751 TCTGAAACAG TTTGGCGAAC CGGACTATGT TGACGATATN TTGAGCGGCG
2801 GTATGTCCGA CGATTGCTG GGAATCAGCC GGAGCGGCGA CGGCGAAACC
2851 GATCCGATGT ACGACGAGGC CGTGTCNGTT GTTTTAAAA CGCGCAAAGC
2901 CAGCATTTCT GCGCTGCAGC GCGCATTGCG TATCGGCTAT AATCGCGCGC
2951 CGCGTCTGAT TGACCAGATG GAGGCGGAAG GCATTGTGTC CGCACCAGAA
3001 CACAACGGCA ACCGTACGAT TCTCGTCCCC TTNGACAATG CTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 492>:

50  
55  
60  
65

```

1 MFIVILVIL LLALAGLFFV RAQSEREWMR EVSAWQEKKG EKQAELEPIK
51 DGMDFPELA LMLFHAVKTA VYWLFGVVVR FCRNYLAHES EPDRFPVPAS
101 ANRADVPTAS DGYSDSGNGT EEAETEEAEA AEEEEADTED IATAVIDNRR
151 IPFDRSIAEG LMPSESEISP VRPVFKEITL EEATRALNSA ALRETKKRYI
201 DAFEKNETAV PKVRVSDTPM EGLQIIGLDD PVLQRTYSRM FDADKEAFSE
251 SADYGFEPYF EKQHPSAFSA VKAENARNAP FRRHAGQGKG QAEAKSPDVS
301 QGQSVSDGTA VRDAXRRVSV NLKEPNKATV SAEARISRLI PESRTVVGKR
351 DVEMPSETEN VFTEKVSSVG YGXPVYDETA DIHIEEPAP wDAWVVEPPEV
401 PKVMPAXDI PPPPVSEIY NRTYEPAGF EQVQSRIAE TDHLADVDLN
451 GGWQEETAAI ANDGSEGVAE RSSGQYLSET EAFGHDSQAV CFFENVPSER
501 PSRRAXDTEA DEGAFOSEET GAVSEHLPTT DLLLPPLFNP GATQTEEXLL
551 XNSITIEEKX AEFKVKVKVV DSYSGPVITR YEIEPDVGVR GNSVLNLEKX
601 LARSLGVASI RVVETILGKT CMGLELPNPK RMIRLSEIF NSPEFAESKS
651 KLTLALGQDI TGQPVVTDLG KAPHLVAGT TSGSKSVGVN AMILSMLFKA
701 APEDVRMIMI DPKMLELSIY EGIPHLAPV VTDMLAANA LNWCVNEMEK
751 RYRLMSFMGV RNLAGXNQKI AEAAARGEKI GNPFSLTPDN PEPLXKLPI
801 VVVVDEFADL MMTAGKKIEE LIARLAQKAR AAGIHLILAT QRPVVDVITG
851 LIKANIPTRI AFQVSSKIDS RTILDQMGAE NLGQGDMLF LPPGTAYPQR
901 VHGAFAFASDEE VHRVVEYLKQ FGEPDYVDDX LSGGMSDDL GISRSGDGET
951 DPMYDEAVSV VLKTRKASIS GVQRALRIGY NRAARLIDQM EAEGIVSAPE
1001 HNGNRTILVP XDNA*

```

ORF58a and ORF58-1 show 96.6% identity in 1014 aa overlap:

		10	20	30	40	50	60
	orf58a.pep	MFWIVLIVILLALLAGLFFVRAQSEREWREVS	SAWQEKKEKQAE	LPEIKDGM	PDFPELA		
5	orf58-1	MFWIVLIVILLALLAGLFFVRAQSEREWREVS	SAWQEKKEKQAE	LPEIKDGM	PDFPELA		
		10	20	30	40	50	60
	orf58a.pep	LMLFHAVKTAVYWL	FVGVRFCRNYLAH	ESPD	RPVP	PASANRADVPT	ASDGYSDSGNGT
10	orf58-1	LMLFHAVKTAVYWL	FVGVRFCRNYLAH	ESPD	RPVP	PASANRADVPT	ASDGYSDSGNGT
		70	80	90	100	110	120
	orf58a.pep	EEAETEEAAEAEAA	ADTEDIATAVIDN	RRIPFDRSIAE	GLMPSESEIS	PVRPVFKEITL	
15	orf58-1	EEAETEEAAEAEAA	ADTEDIATAVIDN	RRIPFDRSIAE	GLMPSESEIS	PVRPVFKEITL	
		130	140	150	160	170	180
	orf58a.pep	EEATRALNSAALRE	TKKRYIDAFEKNE	TAVPKVRVSD	TPMEGLQII	GLDDPVLQRTY	SRM
20	orf58-1	EEATRALNSAALRE	TKKRYIDAFEKNE	TAVPKVRVSD	TPMEGLQII	GLDDPVLQRTY	SRM
		190	200	210	220	230	240
	orf58a.pep	FDADKEAFSE	SADYGFE	PFYFEKQHP	SFAVKAENAR	NAPFRRHAGQ	GKGQAEAKSPDVS
25	orf58-1	FDADKEAFSE	SADYGFE	PFYFEKQHP	SFAVKAENAR	NAPFRRHAGQ	GKGQAEAKSPDVS
		250	260	270	280	290	300
	orf58a.pep	QGQSVSDGTAVR	DAXRRVSVNLKE	PNKATVSAE	ARISRLIPES	SRTVVGKR	DVEMPSETEN
30	orf58-1	QGQSVSDGTAVR	DARRVSVNLKE	PNKATVSAE	ARISRLIPES	SRTVVGKR	DVEMPSETEN
		310	320	330	340	350	360
	orf58a.pep	VFTXVSSVG	YGXPVYDETADI	HIEEPAAPDA	WVVEPPEV	PKVPM	PAXDIPPPPVSEIY
35	orf58-1	VFTXVSSVG	YGXPVYDETADI	HIEEPAAPDA	WVVEPPEV	PKVPM	PAXDIPPPPVSEIY
		370	380	390	400	410	420
	orf58a.pep	NRTYEPPAGFE	QVQRSRIAETD	HLDLADDV	LNGGWQEETA	AIANDGSE	GVAERSSGQYLSET
40	orf58-1	NRTYEPPAGFE	QVQRSRIAETD	HLDLADDV	LNGGWQEETA	AIANDGSE	GVAERSSGQYLSET
		430	440	450	460	470	480
	orf58a.pep	EAFGHDSQAV	CPFENVPSER	PSRRAXDTE	ADGAFQSEET	GAVSEHLPT	TDLLPPLFNP
45	orf58-1	EAFGHDSQAV	CPFENVPSER	PSRRAXDTE	ADGAFQSEET	GAVSEHLPT	TDLLPPLFNP
		490	500	510	520	530	540
	orf58a.pep	GATQTEEXLL	XNSITIEEKXAE	FKVKVVD	SYSGPVITRY	EIEPDV	GVGRNSVLNLEKX
50	orf58-1	GATQTEEXLL	XNSITIEEKXAE	FKVKVVD	SYSGPVITRY	EIEPDV	GVGRNSVLNLEKX
		550	560	570	580	590	600
	orf58a.pep	LARSLGVASIR	VVETILGKTCM	GLELPNPKR	QMIRLSEIF	NSPEFAES	SKSLTLALGQDI
55	orf58-1	LARSLGVASIR	VVETILGKTCM	GLELPNPKR	QMIRLSEIF	NSPEFAES	SKSLTLALGQDI
		610	620	630	640	650	660
	orf58a.pep	LARSLGVASIR	VVETIPGKTCM	GLELPNPKR	QMIRLSEIF	NSPEFAES	SKSLTLALGQDI
60	orf58-1	LARSLGVASIR	VVETIPGKTCM	GLELPNPKR	QMIRLSEIF	NSPEFAES	SKSLTLALGQDI
		670	680	690	700	710	720



5	orf58a.pep	TGQPVVTDLGKAPHLVAGTTGSGKSVGVNAMILSMLFKAAPEDVRMIMIDPKMLELSIY
	orf58-1	TGQPVVTDLGKAPHLVAGTTGSGKSVGVNAMILSMLFKAAPEDVRMIMIDPKMLELSIY
10	orf58a.pep	EGIPHLLAPVVTDMKLAANALNWCVNEMEKRYRLMSFMGVRNLAGXNQKIAEAAARGEKI
	orf58-1	EGIPHLLAPVVTDMKLAANALNWCVNEMEKRYRLMSFMGVRNLAGFNQKIAEAAARGEKI
15	orf58a.pep	GNPFSLTDPNPEPLXKLPFIVVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILAT
	orf58-1	GNPFSLTDPDPEPLEKLPFIVVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILAT
20	orf58a.pep	QRPSVDVITGLIKANIPTRIAFQVSSKIDSRTILDQMGAEENLLGQGDMFLFPGTAYPQR
	orf58-1	QRPSVDVITGLIKANIPTRIAFQVSSKIDSRTILDQMGAEENLLGQGDMFLFLLPGTAYPQR
25	orf58a.pep	VHGAFASDEEVHRVVEYLKQFGEPTYDDXLGGMSDDLGLISRSGDGETDPMYDEAVSV
	orf58-1	VHGAFASDEEVHRVVEYLKQFGEPTYDDILSGGGSEELPGIGRSGDGETDPMYDEAVSV
30	orf58a.pep	VLKTRKASISGVQRALRIGYNRAARLIDQMEAEIGVSAPEHNGNRTILVPLDNAX
	orf58-1	VLKTRKASISGVQRALRIGYNRAARLIDQMEAEIGVSAPEHNGNRTILVPLDNAX

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF58 shows complete identity over a 9aa overlap with a predicted ORF (ORF58ng) from *N. gonorrhoeae*:

40	orf58.pep	ALMLFHAVKTAVYWLFVGVVRFRCRNYLAHESEFDRPVPP	103
	orf58ng	SEPDRPVPPASANRADVPTASDGYSDSGNG	30

The ORF58ng nucleotide sequence <SEQ ID 493> is predicted to encode a protein having partial amino acid sequence <SEQ ID 494>:

45	1	..SEPDRPVPPA	SANRADVPTA	SDGYSDSGNG	TEEAETEAEE	AAAAEADTE
	51	DIATAVIDNR	RIPFDRSIAE	GLMQSESSTS	PVRPVFKEIT	LEEATRALSS
50	101	AALRETKKRY	IDAFEKNGTA	VPKVRVSDTP	MEGLQIIGLD	DPVLQRTYSR
	151	MFDADKEAFS	ESADYGFEPY	FEKQHPSAFS	AVKAENARNA	PFRRHAGQEK
55	201	GQAEAKSPDV	SQGQSVSDGT	AVRDARRRVS	VNLKEPNKAT	VSAEARISRL
	251	IPESRTVVGK	RDVEMPSETE	NVFTETVSSV	GYGGPVYDEA	ADIHIEEPAA
60	301	PDAWVVEPPE	VPEVAVPEID	ILPPPPVSEI	YNRTYEPPAG	FEQAQRSRIA
	351	ETDHLAADVL	NGGWQEETA	IADDGSEGAA	ERSSGQYLSE	TEAFGHDSQA
65	401	VCPFEDVPSE	RPSCRVSDE	ADEGAFQSEE	TGAVSEHLPT	TDLLPLPLFN
	451	PEATQTEEEL	LENSITIEEK	LAEFKVKVKV	VDSYSGPVIT	RYEIEPDVGV
70	501	RGNSVLNLEK	DLARSLGVAS	IRVETIPGK	TCMGLLEPNP	KRQMIRLSEI
	551	FNSPEFAESK	SKLTALGQD	ITGQPVVTDL	GKAPHLVAG	TTGSGKSVGV
75	601	NAMILSMLFK	AAPEDVRMIM	IDPKMLELSI	YEGITHLLAP	VVTDMKLAAN
	651	ALNWCVNEME	KRYRLMSFMG	VRNLAGFNQK	IAEAAARGEK	IGNPFSLTDP
80	701	DPEPLEKLPF	IVVVVDEFAD	LMMTAGKKIE	ELIARLAQKA	RAAGIHLILA
	751	TQRPSVDVIT	GLIKANIPTR	IAFQVSSKID	SRTILDQMG	ENLLGQGDML
85	801	FLPPGTAYPQ	RVHGAFASDE	EVHRVVEYLK	QFGEPTYDD	ILSGGGSEEL
	851	PGIGRSGDGE	TDPMYDEAVS	VVLKTRKASI	SGVQRALRIG	YNRAARLIDQ
90	901	MEAEIGVSAP	EHNGNRTILV	PLDNA*		

This partial gonococcal sequence contains a predicted transmembrane region and a predicted ATP/GTP-binding site motif A (P-loop; double underlined). Furthermore, it has a domain homologous to the FTSK cell division protein of *E. coli*. Alignment of ORF58ng and FtsK (accession number p46889) show a 65 % amino acid identity in 459 overlap:

5	ORF58ng:	467	IEEKLAEFKVKVVDSSYSGPVITRYEIEPDVGVGRNSVLNLEKDLARSLGVASIRVVET	526
			+E +LA+F++K VV+ GPVITR+E+ GV+ + NL +DLARSL ++RVVE	
	FtsK:	868	VEARLADFRKADVVNYSPPGVITRFEFLNLAGPVKAARISNLSRDLARSLSTVAVRVVEV	927
10	ORF58ng:	527	IPGKTCMGLELPNPKRQMIRLSEIFNSPEFAESKSLTLALGQDITGQPVVTDLGKAPHL	586
			IPGK +GLELPN KRQ + L E+ ++ +F ++ S LT+ LG+DI G+PVV DL K PHL	
	FtsK:	928	IPGKPYVGLELPNKKRQTVYLREVLNNAKFRDNPSPLTVVLGKDIAEPVADLAKMPHL	987
15	ORF58ng:	587	LVAGTTGSGKSVGVNAMILSMLFKAAPEDVRMIMIDPKMLELSIYEGITHLLAPVVTDMK	646
			LVAGTTGSGKSVGVNAMILSML+KA PEDVR IMIDPKMLELS+YEGI HLL VVTDMK	
	FtsK:	988	LVAGTTGSGKSVGVNAMILSMLYKAQPEDVRFIMIDPKMLELSYEGIPHLLTEVVTDMDK	1047
20	ORF58ng:	647	LAANALNWCVNEMEKRYRLMSFGVRNLAGFNQKIAEAAARGEKIGNPFSLTPDDPEP--	704
			AANAL WCVNEME+RY+LMS +GVRNLAG+N+KIAEA I +P+ D +	
	FtsK:	1048	DAANALRWCVNEMERRYKLMSALGVRNLAGYNEKIAEADRMMPRIIDPYWKGDSMDAQH	1107
25	ORF58ng:	705	--LEKLPIFVVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILATQRPSVDVITGL	762
			L+K P+IVV+VDEFADLMMT GKK+EELIARLAQKARAAGIHL+LATQRPSVDVITGL	
	FtsK:	1108	PVLKKEPIYIVLVDEFADLMMTVGKKVEELIARLAQKARAAGIHLVLATQRPSVDVITGL	1167
30	ORF58ng:	763	IKANIPTRIAFQVSSKIDSRTILDQMAENLLQGDMFLPFGTAYPQRVHGAFASDEEV	822
			IKANIPTRIAF VSSKIDSRTILDQ GAE+LLG GDML+ P + P RVHGAF D+EV	
	FtsK:	1168	IKANIPTRIAFTVSSKIDSRTILDQGAESLLGMGDMLYSGFNSTLPVRVHGAFVRDQEV	1227
35	ORF58ng:	823	HRVVEYLKQFGPEPDYVDDILSGGGSEELPGIGRSGDGETDPYDEAVSVVLKTRKASISG	882
			H VV+ K G P YVD I S SE G G G E DP++D+AV V + RKASISG	
	FtsK:	1228	HAVVQDWKARGRPQYVDGITSDESEGGAG-GFDGAEELDPLFDQAVQFVTEKRKASISG	1286
40	ORF58ng:	883	VQRALRIGYNRAARLIDQMEAEGIVSAPEHNGNRILVP	921
			VQR RIGYNRAAR+I+QMEA+GIVS HNGNR +L P	
	FtsK:	1287	VQRQFRIGYNRAARIIEQMEAQGIVSEQGHNGNREVLAP	1325

Further work on ORF58ng revealed the complete gonococcal DNA sequence to be <SEQ ID 495>:

1	ATGTTTGGGA	TAGTTTGGAT	CGTTATtgtg	TTGCTTGCGC	TTGCCGGCCT
51	GTTTTTTGTG	CGCGCACAA	CCGAACGCGA	GTGGATGCGC	GAGGTTTCTG
101	CGTGGCAGGA	AAAGAAAGGG	GAAAAACAGG	CGGAGCTGCC	TGAAATCAAA
151	GACGGTATGC	CCGATTTTCC	CGAGTTTTC	CTGATGCTT	TCCATGCCGT
201	CAAAACGGCA	GTGTATTGGC	TGTTTGTCCG	TGTCGTCCGT	TTCTGCCGAA
251	ACTATCTGGC	GCACGAATCC	GAACCGGACA	GGCCCGTTCC	GCCTGCTTCT
301	GCAAACCGTG	CGGATGTTCC	GACCGCATCC	GACGGGTATT	CAGACAGTGG
351	AAACGGGACG	GAAGAAGCGG	AAACGGAAGC	AGCAGAAGCT	GCGGAGGAAG
401	AGGCTGCCgA	TACgGAAGAC	ATTGCAACTG	CCGTAATCGA	CAACCGCCGC
451	ATCCcatTCG	ACCGGAGTAT	TGCTGAAGGG	TTGATGCAGT	CTGAAAGCAA
501	AACTTCGCCC	GTCCGTCCGG	TTTTTAAGGA	AATCACTTTG	GAAGAAGCAA
551	CGCGTGCTTT	AAGCAGCGCG	GCTTTAAGGG	AAACGAAAAA	ACGCTATATC
601	GATGCATTTG	AGAAAAACGG	AACAGCCGTC	CCCAAGTAC	GCGTGTCCGA
651	TACCCGATG	GAAGGGCTGC	AGATTATCGG	TTTGGACGAC	CCTGTGCTTC
701	AACGCACGTA	TTCCCGTATG	TTTGTATGCG	ACAAAGAAGC	GTTTTCGCGAG
751	TCTGCGGATT	ACGGATTTGA	GCCGTATTTT	GAGAAGCAGC	ATCCGTCTGC
801	CTTTTCTGCA	GTCAAAGCCG	AAAATGCACG	GAATGCGCCG	TTCCGCGGTC
851	ATGCAGGGCA	GGAGAAAGGG	CAGGCGGAGG	CAAAATCCCC	GGATGTTTCC
901	CAAGGGCAGT	CCGTTTCAGA	CGGCACAGCC	GTCCCGGATG	CCCGCCGCCG
951	CGTTTCCGTC	AATTTGAAAG	AACCGAACAA	GGCAACGTTT	TCTGCGGAGG
1001	CGCGGATTTT	GCGCTGATT	CCGAAAGTC	GGACGGTTGT	CGGGAACCGG
1051	GATGTCGAAA	TGCCGTCTGA	AACCGAAAAT	GTTTTCACGG	AAACCGTTTC
1101	GTCTGTGGGA	TACGGCGGTC	CGTTTATGA	TGAAGCTGCC	GATATCCATA
1151	TTGAAGAGCC	TGCCGCGCCC	GATGCTTGGG	TGGTCGAACC	ACCCGAAGTG
1201	CCGGAGGTAG	CCGTACCCGA	AATCGATATT	CTGCCGCCGC	CTCCCCGATC
1251	GGAAATCTAC	AACCGTACCT	ATGAGCCGCC	GGCAGGATTC	GAGCAGGCGC
1301	AACGCAGCCG	CATTGCCGAA	ACCGACCATC	TTGCCGCTGA	TGTTTTGAAT

5  
10  
15  
20  
25  
30  
35

```

1351 GGAGGTTGGC AGGAGGAAAC CGCCGCTATT GCAGATGACG GCAGTGAGGG
1401 TGCGGCAGAG CGGTCAAGCG GGCAATATCT GTCGGAAACC GAAGCGTTCG
1451 GGCATGACAG TCAGGCGGTT TGTCGGTTTG AAGATGTGCC GTCTGAACGC
1501 CCGTCCTGCC GGGTATCGGA TACGGAAGCG GATGAAGGGG CGTTCCAATC
1551 GGAAGAGACC GGTGCGGTAT CCGAACACCT GCCGACAACC GACCTGCTTC
1601 TGCCTCCGCT GTTCAATCCC GAGGCGACGC AAACCGAAGA AGAACTGTTG
1651 GAAAACAGCA TCACCATCGA AGAAAAATTG GCGGAGTTCA AAGTCAAGGT
1701 CAAGGTTGTC GATTCTTATT CCGGCCCCGT GATTACGCGT TATGAAATCG
1751 AACCCGATGT CGGCGTGCGC GGCAATTCCG TTCTGAATTT GAAAAAGAC
1801 TTGGCGCGTT CGCTCGGCGT GGCTTCCATC CGCGTTGTGC AAACCATCCC
1851 CGGCAAAACC TGCATGGGTT TGGAACTTCC GAACCCGAAA CGCCAAATGA
1901 TACGCCTGAG CGAAATTTTC AATTGCCCCG AGTTTGCCGA ATCCAAATCC
1951 AAGCTGACGC TCGCGCTCGG TCAGGACATT ACCGGACAGC CCGTCGTAAAC
2001 CGACTTGGGC AAAGCACCGC ATTTGCTGGT TGCCGGCAGC ACCGGTTCGG
2051 GCAAAATCGGT GGGTGTCAAC GCGATGATTC TGTCTATGCT TTCAAAGCC
2101 GCGCCGGAAG ACGTGCGTAT GATTATGATC GATCCGAAAA TGCTGGAATT
2151 GAGCATTTAC GAAGGCATCA CGCACCTGCT CGCCCTGTG GTTACCGATA
2201 TGAAGCTGGC GGCAACCGC CTGAACTGGT GTGTTAACGA AATGGAATAA
2251 CGCTACCGCC TGATGAGCTT TATGGGCGTG CGCAATCTTG CGGGCTTCAA
2301 CCAAAAAATC GCCGAAGCCG CAGCAAGGGG AGAAAAATC GGCAATCCGT
2351 TCAGCCTCAC GCCGACGAT CCCGAACCTT TGGAAAAACT GCCGTTTATC
2401 GTGGTCTGGT TCGATGAGTT TGCCGATTG ATGATGACGG CAGGCAAGAA
2451 AATCGAAGAA CTGATTGCGC GCCTCGCCCA AAAAGCCCCG CCGGCAGCA
2501 TCCACCTTAT CCTTGCCACA CAACGCCCCA GCGTCGATGT CATCACGGGT
2551 CTGATTAAGG CGAACATCCC GACGCGTATC GCGTTCCAAG TGTCAGCAA
2601 AATCGACAGC CGCACGATTC TCGACCAAT GGGCGCGGAA AACCTGCTCG
2651 GTCAGGGCGA TATGCTGTTT CTGCCGCCGG GTACTGCCTA TCCGCAGCGC
2701 GTTACGCGC CGTTTGCCTC GGATGAAGAG GTGCACCGCG TGGTCGAATA
2751 TCTGAAGCAG TTTGGCGAGC CGGACTATGT TGACGATATT TTGAGCGGCG
2801 GCGGCAGCGA AGAGCTGCCG GGCATCGGGC GCAGCGGCGA CGGCGAAACC
2851 GATCCGATGT ACGACGAGGC CGTATCCGTT GTCTGAAAA CGGCGAAAGC
2901 CAGCATTTTC GCGGTACAGC GCGCCTTGC CATCGGCTAC AACCGCGCCG
2951 CGGTCTGAT TGACCAAAATG GAAGCGGAAG GCATTGTGTC CGCACCGGAA
3001 CACACCGCA ACCGTACGAT TCTCGTCCCC TTGACAATG CTTGA

```

35 This corresponds to the amino acid sequence <SEQ ID 496; ORF58ng-1>:

40  
45  
50  
55

```

1 MFWIVLIVIV LLALAGLFFV RAQSEREWMR EVSAWQEKKG EKQAELEPEIK
51 DGMPDFPEFS LMLFHAVKTA VYWLFGVVR FCRNYLAHES EPDRPVPPAS
101 ANRADVPTAS DGYSDSNGNT EEAETEAEEA AEEEAADTED IATAVIDNRR
151 IPFDRSIAEG LMQSESKTSP VRPVFKEITL EEATRALSSA ALRETKKRYI
201 DAFEKNGTAV PKVRVSDTPM EGLQIIGLDD PVLQRTYSRM FDADKEAFSE
251 SADYGFEPYF EKQHPSAFSA VKAENARNAP FRRHAGQKEG QAEAKSPDVS
301 QGQSVSDGTA VRDARRRVSV NLKEPNKATV SAEARISRLI PESRTVVGKR
351 DVEMPSETEN VFTETVSSVG YGGPVYDEAA DIHIEPPAAP DAWVVEPPEV
401 PEVAVPEIDI LPPPVSEIY NRTYEPFAGF EQAQRSRIAE TDHLAADVLN
451 GGWQEETAAI ADDGSEGAAE RSSGQYLSET EAFGHDSQAV CPFEDVPSEK
501 PSCRVSDETA DEGAQFSEET GAVSEHLPTT DLLLPPLFNP EATQTEEELL
551 ENSITIEEKL AEFKVKVKVV DSYSGPVITR YEIEPDVGVR GNSVLNLEKD
601 LARSLGVASI RVVETIPGKT CMGLELPNPK RQMIRLSEIF NSPEFAESKS
651 KLTLALGQDI TGQPVVTDLG KAPHLVAGT TSGKSVGVN AMILSMLFKA
701 APEDVRMIMI DPKMLELSIY EGITHLLAPV VTDMLAANA LNWCVNEMEK
751 RYRILMSFMGV RNLAFGNQKI AEAARGEKI GNPFSLTPDD PEPELEKLPFI
801 VVVVDEFAFL MMTAGKKIEE LIARLAQKAR AAGIHLILAT QRPVSDVITG
851 LIKANIPTRI AFQVSSKIDS RTILDQMGAE NLLGQGDMLF LPPGTAYPQR
901 VHGAFAFASDEE VHRVVEYLKQ FGEPTYVDDI LSGGGSEELP GIGRSGDGET
951 DPMYDEAVSV VLKTRKASIS GVQRALRIGY NRAARLIDQM EAEGIVSAFE
1001 HNGNRTILVP LDNA*

```

ORF58ng-1 and ORF58-1 show 97.2% identity in 1014 aa overlap:

60  
65

```

10 20 30 40 50 60
orf58-1.pep MFWIVLIVILLALAGLFFVRAQSEREWMREVS AWQEKKG EKQAELEPEIKDGMPDFPELA
|||||:|||||
orf58ng-1 MFWIVLIVILLALAGLFFVRAQSEREWMREVS AWQEKKG EKQAELEPEIKDGMPDFPEFS
10 20 30 40 50 60

70 80 90 100 110 120
orf58-1.pep LMLFHAVKTAVYWLFGVVRFCRNYLAHES EPDRPVPPASANRADVPTASDGYSDSNGNT
|||||

```

	orf58ng-1	LMLFHAVKTAVYWLFVGVVFR	CNRYLAHES	EPDRPVPPAS	ANRADVPTAS	DGYSDSGNGT
		70	80	90	100	110 120
5	orf58-1.pep	EEAETEEAEAEAAEAAADTED	IATAVIDNRR	IPFDRSIAEGL	MPSESEISPV	RPVFKEITL
	orf58ng-1	EEAETEEAEAEAAEAAADTED	IATAVIDNRR	IPFDRSIAEGL	MQSESKTSPV	RPVFKEITL
10	orf58-1.pep	EEATRALNSAALRETKKRYID	AFEKNETAV	PKVRVSDTP	MEGLQIIGL	DDPVLQRTYSHM
	orf58ng-1	EEATRALSSAALRETKKRYID	AFEKNGTAV	PKVRVSDTP	MEGLQIIGL	DDPVLQRTYSRM
15	orf58-1.pep	FDADKEAFSESADYGFE	PFYFEKQHPS	AFSAVKAEN	ARNAPFHR	HAGQKGQAEAKSPDVS
	orf58ng-1	FDADKEAFSESADYGFE	PFYFEKQHPS	AFSAVKAEN	ARNAPFRR	HAGQKGQAEAKSPDVS
20	orf58-1.pep	QGQSVSDGTAVRDARRRV	SVNLKEPNK	ATVSAEARI	SRLIPESQ	TVVGKRDVEMPSETEN
	orf58ng-1	QGQSVSDGTAVRDARRRV	SVNLKEPNK	ATVSAEARI	SRLIPESR	TVVGKRDVEMPSETEN
25	orf58-1.pep	VFTETVSSVGYGGPVYDE	TAADIHIEE	PAAPDAWV	VEPPEV	PKVPMTAIDIQPPPVSEIY
	orf58ng-1	VFTETVSSVGYGGPVYDE	TAADIHIEE	PAAPDAWV	VEPPEV	PEVAVPEIDILPPPVSEIY
30	orf58-1.pep	NRTYEPSPSGFEQVQRS	RIAETDHL	ADDVNLGG	WQEETA	AAIADDGSEGA
	orf58ng-1	NRTYEPSPSGFEQVQRS	RIAETDHL	ADDVNLGG	WQEETA	AAIADDGSEGA
35	orf58-1.pep	EAFGHDSQAVCPFEN	VPSERPSCR	VSDTEADE	GAFQSEET	GAVSEHLPTD
	orf58ng-1	EAFGHDSQAVCPFED	VPSERPSCR	VSDTEADE	GAFQSEET	GAVSEHLPTD
40	orf58-1.pep	EATQTEEELENSITIE	EKLAEFKV	KVVD	SYSGPVIT	RYEIEPDV
	orf58ng-1	EATQTEEELENSITIE	EKLAEFKV	KVVD	SYSGPVIT	RYEIEPDV
45	orf58-1.pep	LARSLGVASIRVVETI	PGKTCMGL	ELPNPKRQ	MIRLSEIF	NSPFAESKSKLT
	orf58ng-1	LARSLGVASIRVVETI	PGKTCMGL	ELPNPKRQ	MIRLSEIF	NSPFAESKSKLT
50	orf58-1.pep	TGQPVVTDLGKAPHLL	VAGTTGSG	KSVGVNAMI	SMLFKAAP	EDVRMIMIDPKM
	orf58ng-1	TGQPVVTDLGKAPHLL	VAGTTGSG	KSVGVNAMI	SMLFKAAP	EDVRMIMIDPKM
55	orf58-1.pep	EGIPHLLAPVVTD	MKLAANAL	NWCNEMEK	RYRLMSFM	GVRNLAGFNQKIA
	orf58ng-1	EGITHLLAPVVTD	MKLAANAL	NWCNEMEK	RYRLMSFM	GVRNLAGFNQKIA
60	orf58-1.pep	EGIPHLLAPVVTD	MKLAANAL	NWCNEMEK	RYRLMSFM	GVRNLAGFNQKIA
	orf58ng-1	EGITHLLAPVVTD	MKLAANAL	NWCNEMEK	RYRLMSFM	GVRNLAGFNQKIA
65	orf58-1.pep	EGIPHLLAPVVTD	MKLAANAL	NWCNEMEK	RYRLMSFM	GVRNLAGFNQKIA
	orf58ng-1	EGITHLLAPVVTD	MKLAANAL	NWCNEMEK	RYRLMSFM	GVRNLAGFNQKIA
70	orf58-1.pep	EGIPHLLAPVVTD	MKLAANAL	NWCNEMEK	RYRLMSFM	GVRNLAGFNQKIA
	orf58ng-1	EGITHLLAPVVTD	MKLAANAL	NWCNEMEK	RYRLMSFM	GVRNLAGFNQKIA

5	orf58-1.pep	GNPFSLTPDDPEPLEKLPFIVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILAT				
	orf58ng-1	GNPFSLTPDDPEPLEKLPFIVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILAT	790	800	810	820
10	orf58-1.pep	QRPSVDVITGLIKANIPTRIAFQVSSKIDSRTILDQMGAEENLLGQGDMLFLLPGTAYPQR	850	860	870	880
	orf58ng-1	QRPSVDVITGLIKANIPTRIAFQVSSKIDSRTILDQMGAEENLLGQGDMLFLLPGTAYPQR	850	860	870	880
15	orf58-1.pep	VHGAFASDEEVHRVVEYLKQFGEPTYDDILSGGGSEELPGIGRSGDDETDPMYDEAVSV	910	920	930	940
	orf58ng-1	VHGAFASDEEVHRVVEYLKQFGEPTYDDILSGGGSEELPGIGRSGDDETDPMYDEAVSV	910	920	930	940
20	orf58-1.pep	VLKTRKASISGVQRALRIGYNRAARLIDQMEAEIGVSAPEHNGNRTILVPLDNAX	970	980	990	1000
	orf58ng-1	VLKTRKASISGVQRALRIGYNRAARLIDQMEAEIGVSAPEHNGNRTILVPLDNAX	970	980	990	1000

Furthermore, ORF58ng-1 shows significant homology to the *E.coli* protein FtsK:

25	sp P46889 FTSK_ECOLI CELL DIVISION PROTEIN FTSK >gi 1651412 gnl PID d1015290 (D1 division protein FtsK [Escherichia coli] >gi 1651418 gnl PID d1015296 (D90727) Cell division protein FtsK [Escherichia coli] >gi 1787117 (AE000191) cell division protein FtsK [Escherichia coli] Length = 1329					
	Score = 576 bits (1469), Expect = e-163					
30	Identities = 301/459 (65%), Positives = 353/459 (76%), Gaps = 5/459 (1%)					
	Query: 556	IEEKLAEFKVKVVDVSYSGPVITRYEIEPDVGVGRNSVLNLEKDLARSLGVASIRVVET	615			
35	Sbjct: 868	+E +LA+F++K VV+ GPVITR+E+ GV+ + NL +DLARSL ++RVVE				
		VEARLADFRIKADVNNYSPGPVITRFEINLAPGVKAARISNLSRDLARSLSTVAVRVVEV	927			
40	Query: 616	IPGKTCMGLELPNPKRQMIRLSEIFNSPEFAESKSKLTALGQDITGQPVVTDLGKAPHL	675			
	Sbjct: 928	IPGKPYVGLLEPNKKRQTVYLREVLDNAKFRDNPSPLTVVLGKDIAGEPVVADLAKMPHL	987			
45	Query: 676	LVAGTTGSGKSVGVNAMILSMLFKAAPEDVRMIMIDPKMLELSIYEGITHLLAPVVTDMK	735			
	Sbjct: 988	LVAGTTGSGKSVGVNAMILSMLYKAQPEDVRFIMIDPKMLELSVYEGIPHLLTEVVTDMK	1047			
50	Query: 736	LAANALNWCVNEMEKRYRLMSFMGVRNLAGFNQKIAEAAARGEKIGNPFSLTPDDPEP--	793			
	Sbjct: 1048	AANAL WCVNEME+RY+LMS +GVRNLAG+N+KIAEA I +P+ D +				
55	Query: 794	--LEKLFPFIVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILATQRPSVDVITGL	851			
	Sbjct: 1108	L+K P+IVV+VDEFADLMMT GKK+EELIARLAQKARAAGIHL+LATQRPSVDVITGL	1167			
60	Query: 852	IKANIPTRIAFQVSSKIDSRTILDQMGAEENLLGQGDMLFLLPGTAYPQRVHGAFASDEEV	911			
	Sbjct: 1168	IKANIPTRIAF VSSKIDSRTILDQ GAE+LLG GDML+ P + P RVHGAF D+EV	1227			
65	Query: 912	HRVVEYLKQFGEPTYDDILSGGGSEELPGIGRSGDGETDPMYDEAVSVVLKTRKASISG	971			
	Sbjct: 1228	H VV+ K G P YVD I S SE G G G E DP++D+AV V + RKASISG				
70	Query: 972	VQALRIGYNRAARLIDQMEAEIGVSAPEHNGNRTILVP	1010			
	Sbjct: 1287	VQR RIGYNRAAR+I+QMEA+GIVS HNGNR +L P				
75		VQRQFRIGYNRAARIIEQMEAQIVSEQHNGNREVLP	1325			

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

**Example 59**

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 497>:

```

      1  ATGATTTATC AAAGAAACCT CATCAAAGAA CTCTCTTTTA CCGCCGTCGG
      51  CATTTTCGTC GTCCTCTTGG CGGTATTGGT CTCCACGCAG GCAATCAACC
5      101  TGCTCGGCCG TGCCGCCGAC GGGC..GTGA TCGCCATCGA TGCCGTGTGG
      151  GCATTGGTCG GCTTCTGGGT C.....
          //
      901  .....A TTGCCATCGG TTTGTTTTTA ATTTACAAA ACGGGCTGAC
      951  CCTGCTTTTT GAAGCCGTGG AAGACGGCAA AATCCATTTT TGGCTCGGAC
10     1001  TGCTGCCTAT GCACATTATC ATGTTTGTCC TTGCATCAT CCTGTTGCCG
      1051  GTCCGCAGTA TGCCAGCCA GCCCTTCTGG CAGGCGGTTG GCAAAAGTCT
      1101  GACATTGAAA GCGGAAAT GA

```

This corresponds to the amino acid sequence <SEQ ID 498; ORF101>:

```

      1  MIYQRNLIKE LSFTAVGIFV VLLAVLVSTQ AINLLGRAAD GXVIAIDAVL
15     51  ALVGFVW... ..
          //
      301  ...IAIGLFL IYQNLTLF EAVEDGKIHF WLGLLPMHII MFVLALILLR
      351  VRSMPSQFW QAVGKSLTLK GKK*

```

Further work revealed the complete nucleotide sequence <SEQ ID 499>:

```

20     1  ATGATTTATC AAAGAAACCT CATCAAAGAA CTCTCTTTTA CCGCCGTCGG
      51  CATTTTCGTC GTCCTCTTGG CGGTATTGGT CTCCACGCAG GCAATCAACC
      101  TGCTCGGCCG TGCCGCCGAC GGGCGTGTCC CCATCGATGC CGTGTGGCA
      151  TTGGTCGGCT TCTGGGTGAT CGGTATGACG CCGCTTTTGC TGGTGTGAC
      201  CGCATTTATC AGTACGTTGA CCGTGTGAC CCGCTACTGG CGCGACAGCG
25     251  AAATGTCGGT CTGGCTATCC TGCGGATTGG CATTGAAACA ATGGATACGC
      301  CCGGTGATGC AGTTTGCCGT GCCGTTTGCC GTTTTGTTG CCGTCATGCA
      351  GCTTTGGGGT ATACCGTGGG CAGAGCTACG CAGCCGCGAA TACGCTGAAA
      401  TCCTGAAGCA GAAGCAGGAA TTGTCTTTGG TGGAGGCAGG CGAGTCAAC
      451  AGTTTGGGCA AGCGCAACGG CAGGGTTTAT TTTGTGAAA CCTTCGATAC
30     501  CGAATCCGGC ATCATGAAA ACCTGTTCTT GCGCGAACAG GACAAAAACG
      551  GCGGCGACAA CATCATCTTC GCCAAAGAAG GTAACCTCTC GCTGAACGAC
      601  AACAAACGCA CGCTCGAATT GCGCCACGGC TACCGTTACA GCGGCACGCC
      651  CGGACGCGCC GACTACAATC AGGTTTCCTT CCAAAACTC AACCTGATTA
      701  TCAGCACCAC GCCCAAATC ATCGACCCCG TTTCCACCG CCGTACCATT
35     751  CCGACCGCCC AACTGATTGG CAGCAGCAAC CCGCAACATC AGGCGGAATT
      801  GATGTGGCGC ATCTCGCTGA CCGTCAGCGT CCTCCTACTC TGCCTGCTTG
      851  CCGTGCCGCT TTCCTATTTC AACCCGCGCA GCGGACATAC CTACAATATC
      901  TTGATTGCCA TCGGTTTGT TTTAATTTC CAAAACGGGC TGACCCTGCT
      951  TTTTGAAGCC GTGGAAGACG GCAAAATCCA TTTTGGCTC GGACTGCTGC
40     1001  CTATGCACAT TATCATGTTT GCCGTTGCAC TCATCCTGTT GCGCGTCCGC
      1051  AGTATGCCCA GCCAGCCCTT CTGGCAGGCG GTTGGCAAAA GTCTGACATT
      1101  GAAAGCGGGA AAATGA

```

This corresponds to the amino acid sequence <SEQ ID 500; ORF101-1>:

```

45     1  MIYQRNLIKE LSFTAVGIFV VLLAVLVSTQ AINLLGRAAD GRVAIDAVLA
      51  LVGFVWIGMT PLLLVLTAFI STLTVLTRYW RDSEMSVWLS CGLALKQWIR
      101  PVMQFAVPFA VLVAVMQLWV IPWAEIERSRE YAEILKQKQE LSLVEAGEFN
      151  SLGKRNGRVY FVETFDTESE IMKNLFLREQ DKNNGDNIIF AKEGNFSLND
      201  NKRTLELRHG YRYSCTPGRA DYNQVSFQKL NLIISTPKL IDPVSHRRTI
50     251  PTAQLIGSSN PQHQAEIMWR ISLTVSVLLL CLLAVPLSYF NPSRGHTYNI
      301  LIAIGLFLIY QNLTLTLLFEA VEDGKIHFWL GLLPMHIIME AVALILLRVR
      351  SMPSPFWQA VGRSLTLKGG K*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF101 shows 91.2% identity over a 57aa overlap and 95.7% identity over a 69aa overlap with

55 an ORF (ORF101a) from strain A of *N. meningitidis*:

```

      10      20      30      40      50
orf101.pep MIYQRNLIKELSFTAVGIFVLLAVLVSTQAINLLGRAADGXVIAIDAVLALVGFVWX
|||||
orf101a    MIYQRNLIKELSFTAVGIFVLLAVLVSTQAINLLGXAADXRX-AIDAVLALVGFVWXXM
      10      20      30      40      50
//
      90      100     110
orf101.pep .....IAIGLFLIYQNGLTLLFEAVEDGKIHFWLGL
|||||
orf101a    LTVSVLLLCLLAVPLSYFNPRSGHTYNILXAIGLFLIYQNGLTLLFEAVEDGKIHFWLGL
      280     290     300     310     320     330

      120     130     140     150
orf101.pep LPMHIIMFVLALILLRVRSMPSQPFQAVGKSLTLKGGKX
|||||
orf101a    LPMHIIMFVIAIVLLRVRSMPSQPFQAVGKSLTLKGGKX
      340     350     360     370

```

The complete length ORF101a nucleotide sequence <SEQ ID 501> is:

```

20 1 ATGATTTATC AAAGAAACCT CATCAAAGAA CTCTCTTTTA CCGCCGTCGG
51 CATTTTCGTC GTCCTCTTGG CGGTATTGGT CTCCACGCAG GCAATCAACC
101 TGCTCGGCCN TGCCGCGCAG NGGCGTNTCG CCATCGATGC CGTGTGGCA
151 TTGGTCGGCT TCTGGGTCNN NNGNATGACG CCGCTTTTGC TNGTGTGAC
201 CGCATTTATC AGTACGTTGA CCGTGTGAC CCGCTACTGG CGNGACAGCG
251 AAATGTCGGT CTGGNTATCC TGGCGATTGG CATTGAAACA ATGGATACGC
25 301 CCGGTGATGC AGTTTGCCGT GCCGTTTGCC GTTTTGTTG CCGTCATGCA
351 GCTTTGGGTG ATACCGTGGG CAGAGCTACG CAGCCGCGAA TACGCTGAAA
401 TCCTGAAGCA GAAGCAGGAA TTGTCTTTGG TGGAGGCAGG CGGGTTCAAC
451 AGTTTGGGCA AGCGCAACGG CAGGTTTAT TTGTGCGAAA CCTTCGATAC
501 CGAATCCGGC ATCATGAAAA ACCTGTTTCT GCGCGAACAG GACAAAAACG
30 551 GCGGCGACAA CATCATCTTC NCCAAAGAAA GTAACCTCTC GCTGAACGAC
601 AACAAACGCA CGCTCGAATT GCGCCACGGC TACCGTTACA GCGGCACGCC
651 CGGACGCGCC GACTACAATC AGGTTTCCTT CCNAAAACCT AACCTGATTA
701 TCAGCACCAC GCCCAAATC ATCGACCCCG TTTCCCACCG CCGTACNATN
751 CCNACNGCCC AACTGATTGG CAGCAGCAAC CCGCAACATC ANGCGGAATT
35 801 GATGTGGCGC ATCTCGCTGA CCGTCAGCGT CCTCCTACTC TGCCTGCTTG
851 CCGTGCCGCT TTCTGATTTC AACCCGCGCA GCGGACATAC CTACAATATC
901 TTGANTGCCA TCGGTTTGT TTTAATTTAC CAAAACGGGC TGACCTGCT
951 TTTTGAAGCC GTGGAAGACG GCAAAATCCA TTTTGGCTC GGACTGCTGC
40 1001 CTATGCACAT CATCATGTTC GTCATCGCAA TCGTACTTCT GCGCGTCCG
1051 AGCATGCCCA GCCAGCCCTT CTGGCAGGCG GTTGGCAAAA GTCTGACATT
1101 GAAAGGCGGA AAATGA

```

This encodes a protein having amino acid sequence <SEQ ID 502>:

```

45 1 MIYQRNLIKE LSFTAVGIFV VLLAVLVSTQ AINLLGXAAD XRXAIDAVLA
51 LVGFVWXXMT PLLLVLTAFI STLTVLTRYW RDSEMSVWXS CGLALKQWIR
101 PVMQFAVPFA VLVAVMQLWV IPWAE LRSRE YAEILKQKQE LSLVEAGGFN
151 SLGKRNGRVY FVETFDTEG IMKNLFLREQ DKNNGDN IIF XKESNFSLND
201 NKRTLELRHG YRYSGTPGRA DYNQVSFXKL NLIISTPKL IDPVSHRRTX
251 PTAQLIGSSN PQHXAELMWR ISLTVSVLLL CLLAVPLSYF NPRSHTYNI
301 LXAIGLFLIY QNGLTLLFEA VEDGKIHFWL GLLPMHIMF VIAIVLLRVR
50 351 SMPSPFWQA VGKSLTLKGG K*

```

ORF101a and ORF101-1 show 95.4% identity in 371 aa overlap:

```

55 orf101a.pep MIYQRNLIKELSFTAVGIFVLLAVLVSTQAINLLGXAADXRXAIDAVLALVGFVWXXMT 60
orf101-1      MIYQRNLIKELSFTAVGIFVLLAVLVSTQAINLLGRAADGRVAIDAVLALVGFVWIGMT 60

orf101a.pep   PLLLVLTAFISTLTVLTRYWRDSEMSVWXS CGLALKQWIRPVMQFAVPFAVLVAVMQLWV 120
orf101-1      PLLLVLTAFISTLTVLTRYWRDSEMSVWLS CGLALKQWIRPVMQFAVPFAVLVAVMQLWV 120

60 orf101a.pep   IPWAE LRSREYAEILKQKQELSLVEAGGFNSLGKRNGRVYFVETFDTEGIMKNLFLREQ 180
orf101-1      IPWAE LRSREYAEILKQKQELSLVEAGEFNSLGKRNGRVYFVETFDTEGIMKNLFLREQ 180

```

orf101a.pep DKNGGDNIIIFXKESNFSLNDNKRTLELRHGYRSGTPGRADYNQVSFXKLNLIISTTPKL 240  
 orf101-1 DKNGGDNIIIFAKEGNFSLNDNKRTLELRHGYRSGTPGRADYNQVSFQKLNLIISTTPKL 240  
 5 orf101a.pep IDPVSHRRXTPTAQLIGSSNPQHXAELMWIRISLTVSVLLCLLAVPLSYFNPRSGHTYNI 300  
 orf101-1 IDPVSHRRTIPTAQLIGSSNPQHQAELMWIRISLTVSVLLCLLAVPLSYFNPRSGHTYNI 300  
 10 orf101a.pep LXAIGLFLIYQNGLTLLFEAVEDGKIHFWLGLLPMHIIMFVIAIVLLRVRSMPSQPFWQA 360  
 orf101-1 LIAIGLFLIYQNGLTLLFEAVEDGKIHFWLGLLPMHIIMFAVALILLRVRSMPSQPFWQA 360  
 orf101a.pep VGKSLTLKGGK 371  
 15 orf101-1 VGKSLTLKGGK 371

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF101 shows 96.5 % identity in 57aa overlap at the N-terminal domain and 95.1% identity in 61aa overlap at the C-terminal domain, respectively, with a predicted ORF (ORF101ng) from *N.*

### *gonorrhoeae*:

orf101.pep MIYQRNLIKELSFTAVGIFVLLAVLVSTQAINLLGRAADGXVIAIDAVLALVGFVW 57  
 orf101ng MIYQRNLIKELSFTAVGIFVLLAVLVSTQAINLLGRAADGRV-AIDAVLALVGFVWIGM 59  
 25 //  
 orf101.pep IAIGLFLIYQNGLTLLFEAVEDGKIHFWLG 333  
 orf101ng SLTVSVLLCLLAVPLSYFNPRSGHTYNILIAIGLFLIYQNGLTLLFEAVEDGKIHFWLG 331  
 30 orf101.pep LLPMHIIMFVLALILLRVRSMPSQPFWQAVGKSLTLKGGK 373  
 orf101ng LLPMHIIMFVIAIVLLRVRSMPSQPFWQAVG 362

The ORF101ng nucleotide sequence <SEQ ID 503> is predicted to encode a protein having partial amino acid sequence <SEQ ID 504>:

1 MIYQRNLIKE LSFTAVGIFV VLLAVLVSTQ AINLLGRAAD GRVAIDAVLA  
 51 LVGFVWIGMT PLLVLVTAFI STLTVLTRYW RDSEMSVWLS CGLALKQWIR  
 101 PVMQFAVPEA ILIAVMQLWV IPWAELSRE YAEILKQKE LSLVEAGEFN  
 151 NLGKRNGRVY FVETFDTESG IMKNLFLREQ DKNGGDNIIIF AKEGNFSLKD  
 201 NKRTLELRHG YRYSCTPGRA DYNQVSFQKL NLIISTTPKL IDPVSHRRTI  
 251 STAQLIGSSN PQHQAELMWR ISLTVSVLLL CLLAVPLSYF NPRSGHTYNI  
 301 LIAIGLFLIY QNGLTLLFEA VEDGKIHFWL GLLPMHIIMF VIAIVLLRVR  
 351 SMPSPFWQA VG...

Further work revealed the complete nucleotide sequence <SEQ ID 505>:

45 1 ATGATTTATC AAAGAAACCT CATCAAAGAA CTCTCTTTTA CCGCCGTCGG  
 51 CATTTTCGTC GTCCTCTTGG CGGTGTTGGT GTCCACGCAG GCGATCAACC  
 101 TGCTTGGCCG CGCAGCTGAC GGGCGTGTCG CCATCGATGC CGTGTGGCC  
 151 TTAGTCGGCT TCTGGGTCAT CGGTATGACC CCGCTTTTGC TGGTGTGAC  
 201 CGCATTTCATC AGCAGCTGA CCGTATTGAC CCGCTACTGG CGCGACAGCG  
 251 AAATGTTCGGT CTGGCTATCC TGGCGATTGG CGTTGAAACA GTGGATACGC  
 301 CCCGTCATGC AGTTTGCCGT GCCGTTTGCC ATCCTGATTG CCGTCATGCA  
 351 GCTTTGGGTG ATACCGTGGG CAGAGCTGCG CAGCCGCGAA TATGCCGAAA  
 401 TTTGAAGCA GAAGCAGGAA TTGTCTTTGG TGGAAGCCGG CGAGTTCAAT  
 451 AACTTGGGCA AGCGCAACGG Cagggtttaa Ttcgtcgaa CCTTTGACAC  
 501 CGaatccgGC ATCATGAAAA ACCTGTtcct GcGCGAACAG GACAAAAACG  
 551 gcggcgacaA CATCATCTTC GCcaaaGAag gtaactTctc gctgaaggac  
 601 AACAAAcgca cgctcgaaATT GCGCCACGGC TACCGTTACA GCGGcagcgc  
 651 CGGacGCGCc gactaCAATC AGGTtTcctt cCAAAAcTc aacctgATta  
 701 TCAGCACCAC GCCCAAacTT ATCGaccCCG TTTCCACCCG CCGCACCATT



5  
 751 tgcacCGCCC AAcTGATTGG CAGCAGCAAT CCGCAACATC AGGCAGAATT  
 801 GATGTGGCGC ATCTCGCTGA CCGTCAGCGT CCTCCTGCTC TGCCTACTCG  
 851 CCGTGCCGCT TTCCTATTTC AACCCGCGCA GCGGACATAC CTACAATATC  
 901 TTGATTGCCA TCGGTTTGT TTTAATTAC CAAAACGGGC TGACCCTGCT  
 951 TTTTGAAGCC GTGAAGACG GCAAATCCA TTTTGGGCTC GGACTGCTGC  
 1001 CTATGCACAT CATCATGTTC GTCATCGCAA TCGTACTTCT GCGCGTCCGC  
 1051 AGTATGCCCA GCCAGCCCTT CTGGCAGGCG GTTGGCAAAA GTCTGACATT  
 1101 GAAAGgcgGA AATGA

This corresponds to the amino acid sequence <SEQ ID 506; ORF101ng-1>:

10  
 1 MIYQRNLIKE LSFTAVGIFV VLLAVLVSTQ AINLLGRAAD GRVAIDAVLA  
 51 LVGFVWIGMT PLLLVLTAFI STLTVLTRYW RDSEMSVWLS CGLALKQWIR  
 101 PVMQFAVPFA ILIAVMQLWV IPWAE LRSRE YAEILKQKQE LSLVEAGEFN  
 151 NLGKRNGRVY FVETFDTESE IMKNLFLREQ DKNNGDNIIIF AKEGNFSLKD  
 201 NKRTLELRHG YRYSCTPGRA DYNQVSFQKL NLIISTTPKL IDPVSHRRTI  
 15  
 251 STAQLIGSSN PQHQAE LMRWISLTVSVLLL CLLAVPLSYF NPRSGHTYNI  
 301 LIAIGLFLIY QNGLTLLFEA VEDGKIHFVW LLLPMHIIMF VIAIVLLRVR  
 351 SMPSPFFWQA VGKSLTLKGG K\*

ORF101ng-1 and ORF101-1 show 97.6% identity in 371 aa overlap:

20  
 orf101-1.pep 10 20 30 40 50 60  
 MIYQRNLIKELSFTAVGIFVLLAVLVSTQAINLLGRAADGRVAIDAVLALVGFVWIGMT  
 orf101ng-1 MIYQRNLIKELSFTAVGIFVLLAVLVSTQAINLLGRAADGRVAIDAVLALVGFVWIGMT  
 10 20 30 40 50 60  
 25  
 orf101-1.pep 70 80 90 100 110 120  
 PLLLVLTAFISTLTVLTRYWRDSEMSVWLS CGLALKQWIRPVMQFAVPFAVLAVMQLWV  
 orf101ng-1 PLLLVLTAFISTLTVLTRYWRDSEMSVWLS CGLALKQWIRPVMQFAVPFAVLAVMQLWV  
 70 80 90 100 110 120  
 30  
 orf101-1.pep 130 140 150 160 170 180  
 IPWAE LRSREYAEILKQKQELSLVEAGEFN SLGKRNGRVYFVETFDTESEIMKNLFLREQ  
 orf101ng-1 IPWAE LRSREYAEILKQKQELSLVEAGEFN SLGKRNGRVYFVETFDTESEIMKNLFLREQ  
 130 140 150 160 170 180  
 35  
 orf101-1.pep 190 200 210 220 230 240  
 DKNNGDNIIIFAKEGNFSLNDNKRTLELRHGYRYSCTPGRADYNQVSFQKLNLIISTTPKL  
 orf101ng-1 DKNNGDNIIIFAKEGNFSLNDNKRTLELRHGYRYSCTPGRADYNQVSFQKLNLIISTTPKL  
 190 200 210 220 230 240  
 40  
 orf101-1.pep 250 260 270 280 290 300  
 IDPVSHRRTIPTAQLIGSSNPQHQAELMRWISLTVSVLLLCLLAVPLSYFNPRSGHTYNI  
 orf101ng-1 IDPVSHRRTIPTAQLIGSSNPQHQAELMRWISLTVSVLLLCLLAVPLSYFNPRSGHTYNI  
 250 260 270 280 290 300  
 45  
 orf101-1.pep 310 320 330 340 350 360  
 LIAIGLFLIYQNGLTLLFEAVEDGKIHFVW LLLPMHIIMFAVALILLRVRSMPSQFFWQA  
 orf101ng-1 LIAIGLFLIYQNGLTLLFEAVEDGKIHFVW LLLPMHIIMFVIAIVLLRVRSMPSQFFWQA  
 310 320 330 340 350 360  
 50  
 orf101-1.pep 370  
 VGKSLTLKGGKX  
 orf101ng-1 VGKSLTLKGGKX  
 370

60 Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is

predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 60

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 507>:

```

5       1  ..GGTGGTGGTT TTATCAATGC TTCCTGTGCC ACTTTGACGA CAGCCAAACC
      51  GCAATATCAA GCAGGAGACC TTAGCGCTTT TAAGATAAGG CAAGGCAATG
     101  TTGTAATCGC CGGACACGGT TTGGATGCAC GTGATACCGA TTACACACGT
     151  ATTCTCAGTT ATCATTCCAA AATCGATGCA CCCGTATGGG GACAAGATGT
     201  TCGTGTCTGC GCGGGACAAA ACGATGTGGC CGCAACAGGT GATGCACATT
    10   251  CGCCTATTCT CAATAATGCT GCTGCCAATA CGTCAAACAA TACAGCCAAC
     301  AACGGCACAC ATATCCCTTT ATTTGCGATT GATACAGGCA AATTAGGAGG
     351  TAT.GTATGC CAACAAATC ACCTTGATCA GTACGGTCGA GCAAGCAGGC
     401  ATTCGTAA
  
```

This corresponds to the amino acid sequence <SEQ ID 508; ORF113>:

```

15      1  ..GGGFINASCA TLTTAKPQYQ AGDLSAFKIR QGNVVIAGHG LDARDTDYTR
      51  ILSYHSKIDA PVWGQDVRV AGQNDVAATG DAHSPIILNNA AANTSNNAN
     101  NGTHIPLFAI DTGKLGXVC QQNHLQYGR ASRHS*
  
```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with with *pspA* putative secreted protein of *N.meningitidis* (accession AF030941)

20 ORF and *pspA* show 44% aa identity in 179aa overlap:

```

25      orf113  GGGFINASCATLTAKPQYQAGDLSAFKIRQGNVVIAGHGLDARDTDYTRILSYHSKIDA 60
      pspA      GGG INA+ TLT+ P   G+L+ F + G VVI G GLD D DYTRILS ++I+A
      pspA      GGGLINAASVTLTSGVPVLNNGNLTGFDVSSGKVVIGGKGLDTSADYTRILSRAAEINA 256

      orf113  PVWGQDVRVAGQNDVAATGDAHSPIILXXXXXXXXXXXXXGTHIPLFAIDTGKLGGMYA 120
      pspA      VWG+DV+VV+G+N + G                               + P AIDT LGGMYA
      pspA      GVWGKDVKVVS GKNKLD FDG-----SLAKTASAPSSSDSVTPTVAIDTATLGGMYA 307

      orf113  NKITLISTVEQAGIRNQGWFFASAGNVAVNAEGKLVNTGMIAATGENHAVSLHARNVHN 179
      pspA      +KITLIST A IRN+G+ FA+ G V ++A+GKL N+G I A      +++ A+ V N
      pspA      DKITLISTDNGAVIRNKGRIFAATGGVTLSDGKLSNSGSIDAA----EITISAQTVDN 362
  
```

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF113 shows 86.5% identity in 52aa overlap at the N- terminal part and 94.1% identity in 17aa overlap at the C-terminal part with a predicted ORF (ORF113ng) from *N. gonorrhoeae*:

```

35      orf113                                     GGGFINASCATLTAKPQYQAGDLSAFKIR 30
      orf113ng  SHPSQLNGYIEVGRRAEVVIANPAGIAVNGGGFINASRATLTGQPQYQAGDFSGFKIR 224
      orf113  QGNVVIAGHGLDARDTDYTRILSYHSKIDAPVWGQDVRVAGQNDVAATGDAHSPIILNNA 90
      orf113ng  QGNAVIAGHGLDARDTDFTRILVCQQNHLQYGRGTSRHS 263
      orf113  IDTGKLGXVCQQNHLQYGRASRHS 135
      orf113ng  DFGSGFKIRQGNVVIAGHGLDARDTDFTRILVCQQNHLQYGRGTSRHS 263
  
```

The complete length ORF113ng nucleotide sequence <SEQ ID 509> is predicted to encode a protein having amino acid sequence <SEQ ID 510>:

1 MNKTLRYRVF NRKRGAVVAV AETTKREGKS CADSGSGSVY VKSVSFIPTH  
 51 SKAFCFSSALG FSLCLALGTV NIAFADGIIT DKAAPKTQQA TILQTGNGIP  
 101 QVNIQTPTSA GVSVNQYQAF DVGNRGAILN NSRSNTQTQL GGWIQGNPWL  
 151 TRGEARVVVN QINSSHPSQL NGYIEVGRR AEVVIANPAG IAVNGGGFIN  
 5 ASRATLTGQ PQYQAGDFSG FKIRQGNVI AGHGLDARDT DFTRILVCCQ  
 251 NHLDQYGRTS RHS\*

Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 61

10 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 511>:

1 .TCAACGGGAC ATAGCGAACA AAATTACACT TTGCCGCGAG AAATCACACG  
 51 CAACATTTCA CTGGGTTTCAT TTGCCTATGA ATCGCATCGC AAAGCATTAA  
 101 GCCATCATGC GCCAGCCAA GGCACGTAGT TGCCGCAAAG CAACGGTATT  
 151 TCGCTACCCT ATACGTCCAA TTCTTTTACC CCATTACCCA GCAGCAGCTT  
 201 ATACATTATC AATCCTGTCA ATAAAGGCTA TCTTGTGTAA ACCGATCCAC  
 251 GCTTTGCCAA CTACCGTCAA TGGTTGGGTA GTGACTATAT GCTGGACAGC  
 301 CTCAAACTAG ACCCAAACAA TTTACATAAA CGTTTGGGTG ATGGTTATTA  
 351 CGAGCAACGT TTAATCAATG AACAAATCGC AGAGCTGACA GGGCATCGTC  
 401 GTTTAGACGG TTATCAAAAC GACGAAGAAC AATTTAAAGC CTTAATGGAT  
 20 AATGGCGCGA CTGCGGCACG TTcGATGAAT CTCAGCGTTG GCATTGCATT  
 501 AAGTGCCGAG CAAGTAGCGC AACTGACCAG CGATATTGTT TGGTTGGTAC  
 551 AAAAAGAAGT TAAGCTTCCT GATGGCGGCA CACAAACCGT ATTGGTGCCA  
 601 CAGGTTTATG TACGCGTTAA AAATGGCGAC ATAGACGGTA AAGGTGCATT  
 651 GTTGTCAGGC AGCAATACAC AAATCAATGT TTCAGGCAGC CTGAAAAACT  
 25 CAGGCACGAT TCGAGGgCGC AATGCGCTTA TTATCAATAC CGATACGCTA  
 751 GACAATATCG GTGGGCGTAT TCATGCGCAA AAATCAGCGG TTACGGCCAC  
 801 ACAAGACATC AATAATATTG GCGGCATGCT TTCTGCCGAA CAGACATTAT  
 851 TGCTCAACGC AGGCAACAAC ATCAACAGCC AAAGCACCAC CGCCAGCAGT  
 901 CAAAATACAC AAGGCAGCAG CACCTACCTA GACCGAATGG CAGGTATTTA  
 30 TATCACAGGC AAAGAAAAAG GTGTTT..

This corresponds to the amino acid sequence <SEQ ID 512; ORF115>:

1 .STGHSEQNYT LPREITRNI LGSFAYESHR KALSHHAPSQ GTELPQSNIGI  
 51 SLPYTSNSFT PLPSSSLYII NPVNKGylVE TDPRFANYRQ WLGS DYMLDS  
 101 LKLDPNNLHK RLG DGYEQR LINEQIAELT GHRRLDGYQN DEEQFKALMD  
 35 NGATAARSMN LSVGIALSAE QVAQLTSDIV WLQKEVKLP DGGTQTVLVP  
 201 QVYVRVKNQD IDKGALLSG SNTQINVSGS LKNSGTIAGR NALIINTDTL  
 251 DNIGGRIHAQ KSAVTATQDI NNIGGMLSAE QTL LNAGNN INSQSTTASS  
 301 QNTQGSSTYL DRMAGIYITG KEKGV..

Computer analysis of this amino acid sequence gave the following results:

40 Homology with the pspA putative secreted protein of *N.meningitidis* (accession number AF030941)

ORF115 and pspA protein show 50% aa identity in 325aa overlap:

Orf115: 1 STGHSEQNYTLPREITRNI LGSFAYESHRKALSHHAPSQGTLPQSNIGISLPYTSNSFT 60  
 STG+S Y E++ +I +G AY+ + + P + NGI +T  
 45 pspA: 778 STGYSRSPYEPAPEVS-SIRMGISAYKGYAPQQASDIPGTVVPVVAENGIHPTFT----- 831  
 Orf115: 61 PLPSSSLYIINPVNKGylVETDPRFANYRQWLGS DYMLDSLKLDPNNLHKRLGDGYEQR 120  
 LP+SSL+ I P NKGyl+ETDP F +YR+WLGS YML +L+ DPN++HKRLGDGYEQ+  
 pspA: 832 -LPNSSLFAIAPNNKGylIETDPAFTDYRKWLGSYMLAALQQDPNHIHKRLGDGYEQK 890  
 50 Orf115: 121 LINEQIAELTGHRRLDGYQNDDEEQFKALMDNGATAARSMNLSVGIALSAEQVAQLTSDIV 180  
 L+NEQIA+LTG+RRLDGY NDEEQFKALMDNG T A+ + L+ GIALSAEQVA+LTSDIV  
 pspA: 891 LVNEQIAKLTGYRRLDGYTNDDEEQFKALMDNGITIAKELQLTPGIALSAEQVARLTSDIV 950

Orf115: 181 WL VQKEVKLPDGGTQTVLVPQVYVRVKNQDIDGKALLSGSNTQINVSGSLKN-SGTIAG 239  
 WL + V LPDG TQTVL P+VYVR + D++G+GALLSGS I SG+++N G IAG  
 pspA: 951 WLENETVTLPDGTTQTVLKPQVYVRARPKDMNGQALLSGSVVDIG-SGAIENRGGLIAG 1009

5 Orf115: 240 RNALIINTDTLDNIGGRIHAQKSAVTATQDINNIGGMLSAEQTLNAGXXXXXXXXXXXX 299  
 R ALI+N + N+ G + + A DI N G + AE LLL A  
 pspA: 1010 REALILNAQNIKNLQGDLOGKNI FFAAGSDITNTGS-IGAENALLKASNNIESRSETRS 1068

10 Orf115: 300 XXXXXXXXXXXYLD RMAGIYITGKEKG 324  
 + R+AGIY+TG++ G  
 pspA: 1069 NQNEQGSVRNIGRVAGIYLTGRQNG 1093

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF115 shows 91.9% identity over a 334aa overlap with a predicted ORF (ORF115ng) from

#### 15 *N.gonorrhoeae*:

orf115.pep	STGHSEQNYTLPREITRNISLGSFAYESHK	31
orf115ng	NEQTFGEKKVFSENGKLHNYWRARRKGHDETGHREQNYTLPEEITRDISLGSFAYESHK	71
20 orf115.pep	ALSHHAPSQGTLPQSN-----GISLPYTSNSFTPLPSSSLYIINPVNKGYLVE	81
orf115ng	ALSRHAPSQGTLPQSNRDNIRTAKSNGISLPYTPNSFTPLPGSSLYIINPANKGYLVET	131
25 orf115.pep	DPRFANYRQWLGS DYMLDSLKLDPNNLHKRLGDGYYEQRLINEQIAELTGHRRLDGYQND	141
orf115ng	DPRFANYRQWLGS DYMLGSLKLDPNNLHKRLGDGYYEQRLINEQIAELTGHRRLDGYQND	191
orf115.pep	EEQFKALMDNGATAARSMNLSVGIALSAEQVAQLTSDIVWL VQKEVKLPDGGTQTVLVPQ	201
30 orf115ng	EEQFKALMDNGATAARSMNLSVGIALSAEQAAQLTSDIVWL VQKEVKLPDGGTQTVLMPQ	251
orf115.pep	VYVRVKNQDIDGKALLSGSNTQINVSGSLKNSGTIAGRNALIINTDTLDNIGGRIHAQK	261
35 orf115ng	VYVRVKNGGIDGKALLSGSNTQINVSGSLKNSGTIAGRNALIINTDTLDNIGGRIHAQK	311
orf115.pep	SAVTATQDINNIGGMLSAEQTLNAGNNINSQSTTASSQNTQGSSTYLDRMAGIYITGK	321
orf115ng	SAVTATQDINNIGGILSAEQTLNAGNNINNQSTAKSSQNAQGSSTYLDRMAGIYITGK	371
40 orf115.pep	EKGV	325
orf115ng	EKGVLAAQAGKDINIIAGQISNQSDQGQTRLQAGRDINLDTVQTGKYQEIHFADNHTIR	431

An ORF115ng nucleotide sequence <SEQ ID 513> was predicted to encode a protein having amino acid sequence <SEQ ID 514>:

45	1	MLVQTEKDGL	HNEQTFGEKK	VFSENGKLHN	YWRARRKGHD	ETGHREQNYT
	51	LPPEITRDIS	LGSFAYESH	KALSRHAPSQ	GTELPQSNRD	NIRTAKSNGI
	101	SLPYTPNSFT	PLPGSSLYII	NPANKGYLVE	TDPRFANYRQ	WLGS DYMLGS
	151	LKLDPNLHK	RLGDGYYEQ	LINEQIAELT	GHRRLDGYQN	DEEQFKALMD
	201	NGATAARSMN	LSVGIALSAE	QAAQLTSDIV	WL VQKEVKLP	DGGTQTVLMP
50	251	QVYVRVKNNG	IDGKALLSG	SNTQINVSGS	LKNSGTIAGR	NALIINTDTL
	301	DNIGGRIHAQ	KSAVTATQDI	NNIGGILSAE	QTLNAGNN	INNQSTAKSS
	351	QNAQGSSTYL	DRMAGIYITG	KEKGVLAQA	GKDINIIAGQ	ISNQSDQGQT
	401	RLQAGRDINL	DTVQTGKYQE	IHFADNHTI	RGSTNEVGSS	IQTGKDVTLL
	451	SGNNLNAKAA	EVGSAKGTLA	VYAKNDITIS	SGIHAGQVDD	ASKHTGRSGG
55	501	GNKLVIDKA	QSHHETAQSS	TFEGKQVVLQ	AGNDANILGS	NVISDNTRI
	551	QAGNHVRIGT	TQTQSQSEY	HQTQKSGLMS	AGIGFTIGSK	TNTQENQSQS
	601	NEHTGSTVGS	LKGDITIVAS	KHYEQTGSNV	SSPEGNNLIS	TQSMIDIGAAQ
	651	NQLNSKTTQT	YEQKGLTVAF	SSPVDLAQQ	AI AVAHKAAK	QFDKAKTTAL
	701	MPWRLPMQVG	RLFKQAKAPK	K*		

60 Further work revealed the following partial gonococcal DNA sequence <SEQ ID 515>:

1 TTGCTTGTGC AAACAGAAAA AGACGGTTTG CATAACGAGC AAACCTTTGG  
 51 CGAGAAGAAA GTCTTCAGCG AAAATGGTAA GTTGACAAAC TACTGGCGTG  
 101 CGCGTCGTAA AGGACATGAT GAAACAGGGC ATCGTGAACA AAATTATACT  
 151 TTGCCGGAGG AAATCACACG CGACATTTC A CTGGGTTTCAT TTGCCTATGA  
 201 ATCGCATAGC AAAGCATTA G GCGTCATGC GCCCAGCCAA GGCACGTAGT  
 251 TGCCACAAAG TAACCGGGAT AATATCCGTA CTGCGAAAAG CAACGGTATT  
 301 TCGCTACCTT ATACGCCCAA TTCTTTTACC CCATTACCCG GCAGCAGCTT  
 351 ATACATTATC AATCTGCCA ATAAAGGCTA TCTGTGTGAA ACCGATCCAC  
 401 GCTTTGCCAA CTACCGTCAA TGGTTGGGTA GTGACTATAT GCTGGGCAGC  
 10 451 CTCAAAC TAG ACCCAAACAA TTTACATAAA CGTTTGGGTG ATGGTTATTA  
 501 CGAGCAACGT TTAATCAATG AACAAATCGC AGAGCTGACA GGGCATCGTC  
 551 GTTAGACGG TTATCAAAAC GACGAAGAAC AATTTAAAGC CTTAATGGAT  
 601 AATGGCGCGA CTGCGGCACG TTCGATGAAT CTCAGCGTTG GCATTGCATT  
 651 AAGTGCCGAG CAAGCAGCGC AACTGACCAG CGATATTGTT TGGTTGGTAC  
 15 701 AAAAAGAAGT TAACTTCCT GATGGCGGCA CACAAACCGT ATTGATGCCA  
 751 CAGGTTTATG TACGCGTTAA AAATGGCGGC ATAGACGGTA AAGGTGCATT  
 801 GTTGTCAGGC AGCAATACAC AAATCAATGT TTCAGGCAGC CTGAAAAACT  
 851 CAGGCACGAT TGCAGGGCGC AATGCGCTTA TTATCAATAC CGATACGCTA  
 901 GACAATATCG GTGGCGTAT TCATGCGCAA AAATCAGCGG TTACGGGCCA  
 20 951 ACAAGACATC AATAATATTG GCGGCATTCT TTCTGCCGAA CAGACATTAT  
 1001 TGCTCAATGC GGGTAACAAC ATCAACAACC AAAGCACGGC CAAGAGCAGT  
 1051 CAAAATGCAC AAGGTAGCAG CACCTACCTA GACCGAATGG CAGGTATTTA  
 1101 TATCACAGCG AAAGAAAAAG GTGTTTTAGC AGCGCAGGCA GGCAAGAGCA  
 1151 TCAACATCAT TGCCGGTCAA ATCAGCAATC AATCAGATCA AGGGCAAACC  
 25 1201 CGGCTGCAGG CAGGACGCGA CATTAACCTG GATACGGTAC AAACCGGCCA  
 1251 ATATCAAGAA ATCCATTTTG ATGCCGATAA CCATACCATC CGAGGTTCAG  
 1301 CGAACGAAGT CCGCAGCAGC ATTCAAACAA AAGGCGATGT TACCCtattG  
 1351 TCAGGGAATA ATCTCAATGC CAAAGCTGCC GAAGTCGGCA GCGCAAAAGG  
 1401 CACACTTGCC GTGTATGCTA AAAATGACAT TACTATCAGC TCAGGCATCC  
 30 1451 ATGCCGGCCA AGTTGATGAT GCGTCCAAAC ATACAGGCAG AAGCGCGCGC  
 1501 GGTAATAAAT TAGTCATTAC CGATAAAGCC CAAAGTCATC ACGAAACTGC  
 1551 TCAAAGCAGC ACCTTTGAAG GCAAGCAAGT TGTATTGCAG GCAGGAAACG  
 1601 ATGCCAACAT CCTTGGCAGT AATGTTATTT CCGATAATGG CACCCGGATT  
 35 1651 CAAGCAGGCA ATCATGTTTC CATTGGTACA ACCCAAACCT AAAGCCAAAG  
 1701 CGAAACCTAT CATCAAACCC AAAAATCAGG ATTGATGAGT GCAGGTATCG  
 1751 GCTTCACTAT TGGCAGCAAG ACAAAACAC AAGAAAACCA ATCCCAAAGC  
 1801 AACGAACATA CAGGCAGTAC CGTAGGCAGC CTGAAAGGCG ATACCACCAT  
 1851 TGTGCAAGC AAACACTACG AACAAACCGG CAGCAACGTT TCCAGCCCTG  
 40 1901 AGGCAACAA CCTTATCAGC ACGCAAAGTA TGGATATTGG CGCAGCACAA  
 1951 AACCAATTAA ACAGCAAAAC CACCCAAACC TACGAACAAA AAGGCTTAAC  
 2001 GGTGGCATTC AGTTCGCCCG TTACCGATT TGGCACAACA GCGATTGCCG  
 2051 TAGCACACAA AGCAGCAAA AAGTCGGACA AAGCAAAAC GACCGCGTTA  
 2101 ATGCCATGGC GGCTGCCAAT GCAGGTTGGC AGGCCTATCA AACAGGCAAA  
 2151 GCGCACAAA ACTTAG

45 This corresponds to the amino acid sequence <SEQ ID 516; ORF115ng-1>:

1 LLVQTEKDGL HNEQTFGEKK VFSENGKLHN YWRARRKGHD ETGHREQNYT  
 51 LPEEITRDIS LGSFAYESHK KALSRHAPSQ GTELPQSNRD NIRTAKSNGI  
 101 SLPYTPNSFT PLPGSSLYII NPANKGYLVE TDPRFANYRQ WLGS DYMLGS  
 151 LKLDPNLHK RLGDGYEQR LINEQIAELT GHRRLDGYQN DEEQFKALMD  
 201 NGATAARSMN LSVGIALSAE QAAQLTSDIV WLQKEVKLP DGGTQTVLMP  
 251 QVYVRVKNNG IDGKGALLSG SNTQINVSGS LKNSGTIAGR NALIINTDTL  
 301 DNIGGRIHAQ KSAVTATQDI NNIGGILSAE QTLNLLNAGNN INNQSTAKSS  
 351 QNAQGSSTYL DRMAGIYITG KEKGVLAQA GKDINIIAGQ ISNQSDQGGT  
 401 RLQAGRDINL DTVQTGKYQE IHFDADNHTI RGSTNEVGSS IQTKGDVTL  
 55 451 SGNLNAKAA EVGSAKGTLA VYAKNDITIS SGIHAGQVDD ASKHTGRSGG  
 501 GNKLVIITDKA QSHHETAQSS TFEKGQVVLQ AGNDANILGS NVISDNGTRI  
 551 QAGNHVRIGT TQTQSQSEY HQTKSGLMS AGIGFTIGSK TNTQENQSQS  
 601 NEHTGSTVGS LKGDITIVAS KHYEQTGSNV SSPEGNNLIS TQSM DIGAAQ  
 651 NQLNSKTTQT YEQKGLTVAF SSPVTDLAQQ AIAVAHKAAN KSDAKKTAL  
 60 701 MPWRLPMQVG RPIKQAKAHK T\*

This gonococcal protein (ORF115ng-1) shows 91.9% identity with ORF115 over 334aa:

20 30 40 50 60 70  
 orf115ng-1.p NEQTFGEKKVFSENGKLHNYWRARRKGHDETGHREQNYTLPEEITRDISLGSFAYESHK  
 65 orf115 ||| |||||:||||:||||||| |  
 STGHSEQNYTLPREITRNISLGSFAYESHK  
 10 20 30

5 orf115ng-1.p 80 90 100 110 120 130  
 ALSRHAPSQGTLPQSNRDNIRTAKSNGISLPYTPNSFTPLPGSSLYIINPANKGYLVET  
 orf115 ALSHHAPSQGTLPQSN-----GISLPYTPNSFTPLPSSSLYIINPVNKGKLVET  
 40 50 60 70 80

10 orf115ng-1.p 140 150 160 170 180 190  
 DPRFANYRQWLGS DYMLGSLKLD PNNLHKRLGDGYYEQRLINEQIAELTGHRRLDGYQND  
 orf115 DPRFANYRQWLGS DYMLDSLKLD PNNLHKRLGDGYYEQRLINEQIAELTGHRRLDGYQND  
 90 100 110 120 130 140

15 orf115ng-1.p 200 210 220 230 240 250  
 EEQFKALMDNGATAARSMNLSVGIALSAEQAAQLTSDIVVLVQKEVKLPDGGTQTVLMPQ  
 orf115 EEQFKALMDNGATAARSMNLSVGIALSAEQVAQLTSDIVVLVQKEVKLPDGGTQTVLVPQ  
 150 160 170 180 190 200

20 orf115ng-1.p 260 270 280 290 300 310  
 VYVRVKNGGIDGKGALLSGSNTQINVSGSLKNSGTIAGRNALIINTDTLDNIGGRIHAQK  
 orf115 VYVRVKNGGIDGKGALLSGSNTQINVSGSLKNSGTIAGRNALIINTDTLDNIGGRIHAQK  
 210 220 230 240 250 260

25 orf115ng-1.p 320 330 340 350 360 370  
 SAVTATQDINNIGGILSAEQTLNAGNNINNQSTAKSSQNAQGSSTYLDRMAGIYITGK  
 orf115 SAVTATQDINNIGGMLSAEQTLNAGNNINSQSTTASSQNTQGSSTYLDRMAGIYITGK  
 270 280 290 300 310 320

30 orf115ng-1.p 380 390 400 410 420 430  
 EKGVLAAQAGKDINIAGQISNQSDQGQTRLQAGRDINLDTVQTKGYQEIHFADNHTIR  
 orf115 EKGV

In addition, it shows homology with a secreted *N.meningitidis* protein in the database:

gi|2623258 (AF030941) putative secreted protein [Neisseria meningitidis] Length = 2273

Score = 604 bits (1541), Expect = e-172

Identities = 325/678 (47%), Positives = 449/678 (65%), Gaps = 22/678 (3%)

Query: 1 LLVQTEKDGLHNEQTFGKEKKVSENGKLHNYWRARRKGHDETGHREQNYTLPEEITRDIS 60  
 L+V T + L N++T G K + ++ G L H Y R +KG D TG+ Y E++ I  
 Sbjct: 739 LIVGTPESALDNDDELTKTKTI-TDKGDLHRYHRHHKKGRDSTGYSRSPYEPAPEVS-SIR 796

Query: 61 LGSFAYESHKALSRHAPSQGTLPQSNRDNIRTAKSNGISLPYTPNSFTPLPGSSLYII 120  
 +G AY+ + AP Q +++P + + NGI +T LP SSL+ I  
 Sbjct: 797 MGISAYKGY-----APQQASDIPGTV---VPVVAENGIHPTFT-----LPNSSLFAI 840

Query: 121 NPANKGYLVETDPRFANYRQWLGS DYMLGSLKLD PNNLHKRLGDGYYEQRLINEQIAELT 180  
 P NKGYL+ETDP F +YR+WLGS YML +L+ DPN++HKRLGDGYYEQ+L+NEQIA+LT  
 Sbjct: 841 APNNKGYLIETDPAFTDYRKWLGS GYMLAALQQDPNHIHKRLGDGYYEQKLVNEQIAKLT 900

Query: 181 GHRRLDGYQNDDEEQFKALMDNGATAARSMNLSVGIALSAEQAAQLTSDIVVLVQKEVKLP 240  
 G+RRLDGY NDEEQFKALMDNG T A+ + L+ GIALSAEQ A+LTSDIVWL + V LP  
 Sbjct: 901 GYRRLDGYTNDEEQFKALMDNGITIAKELQLTGIALSAEQVARLTSDIVWLENETVTLTP 960

Query: 241 DGGTQTVLMPQVYVRVKNGGIDGKGALLSGSNTQINVSGSLKN-SGTIAGRNALIINTDT 299  
 DG TQTVL P+VYVR + ++G+GALLSGS I SG+++N G IAGR ALI+N  
 Sbjct: 961 DGTQTQTVLKPQVYVRARPKDMNGQGALLSGSVVDIG-SGAIENRGGLIAGREALILNAQN 1019

Query: 300 LDNIGGRIHAQKSAVTATQDINNIGGILSAEQTLNAGNNINNQSTAKSSQNAQGSSTY 359  
 + N+ G + + A DI N G I AE LLL A NNI ++S +S+QN QGS  
 Sbjct: 1020 IKNLQGDLOGKNIFAAAGSDITNTGSI-GAENALLKASNNIESRSETRSNQNEQGSVRN 1078

Query: 360 LDRMAGIYITGKEKGVLAQAQAGKDINIAGQISNQSDQGQTRLQAGRDINLDTVQTKGYQ 419  
 + R+AGIY+TG++ G + AG +I + A +++NQS+ GQT L AG DI DT + Q  
 Sbjct: 1079 IGRVAGIYLTGRQNGSVLLDAGNNIVLTASELTNQSEDGQTVLNAGGDIRSDTTGISRNQ 1138

5 Query: 420 EIHFDADNHTIRGSTNEVGSSIQTKGDVTLSSGNNLNAKAAEVGSAKGT LAVYAKNDITI 479  
FD+DN+ IR NEVGSI+T+G+++L + ++ +AAEVGS +G L + A DI +  
Sbjct: 1139 NTIFDSDNYVIRKEQNEVGSTIRGNLSLNAKGDIRIRAAEVGSEQRLKLAAGRDIKV 1198

10 Query: 480 SSGIHAGQVDDASKHTGRSGGGNKLVIDKAQSHHETAQSSTFEGKQVVVLQAGNDANILG 539  
+G + +DA K+TGRSGGG K +T ++ + A S T +GK+++L +G D + G  
Sbjct: 1199 EAGKAHTETEDALKYTGRSGGGIKQKMRHLKNQNGQAVSGTLDGKEIILVSGRDITVTG 1258

15 Query: 540 SNVISDNGTRIQAQGNHVRIGTTQTQSQSEYHQTQKSGLM-SAGIGFTIGSKTNTQENQS 598  
SN+I+DN T + A N++ + +T+S+S ++ +KSGLM S GIGFT GSK +TQ N+S  
Sbjct: 1259 SNIIADNHTILSAKNNIVLKAETR SRSAEMNKKEKSGLMGSGGIGFTAGSKKDTQTNRS 1318

20 Query: 599 QSNHTGSTVGSGLKGDTTIVASKHYEQTGSNVSSPEGNNLISTQSM DIGAAQNQLNSKTT 658  
++ HT S VGSL G+T I A KHY QTGS +SSP+G+ IS+ + I AAQN+ + ++  
Sbjct: 1319 ETVSHTESVVGSLNGNTLISAGKHYTQTGSTISSPQGDVGISSGKISIDAAQNRYSEQESK 1378

Query: 659 QTYEQKGLTVAFSSPVT 676  
Q YEQKG+TVA S PV +  
Sbjct: 1379 QVYEQKGVTV AISVPV 1396

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 62

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 517>:

25 1 ..TCAGGGAATA ACCTCAATGC CAAAGCTGCC GAAGTCAGCA GCGCAAACGG  
51 TACACTCGCT GTGTCTGCCA ATAATGACAT CAACATCAGC GCAGGCATCA  
101 ACACGACCCA TGTGTATGAT GCGTCCAAAC ACACAGGCAG AAGCGGTGGT  
151 GGCAATAAAT TAGTCATTAC CGATAAAGCC CAAAGTCATC ACGAAACCGC  
201 CCAAAGCAGC ACCTTTGAAG GCAAGCAAGT TGTATTGCAG GCAGGAAACG  
30 251 ATGCCAACAT CTTTGGCAGC AATGTTATTT CCGATAATGG CACCCAGATT  
301 CAAGCAGGCA ATCATGTTCTG CATTGGTACA ACCCAAACCTC AAAGCCAAAG  
351 CGAAACCTAT CATCAAACCC AGAAATCAGG ATTGATGAGT GCAGGTATCG  
401 GCTTCACTAT TGGCAGCAAG ACAACACAC AAGAAAACCA ATCCCAAAGC  
451 AACGAACATA CAGGCAGTAC CGTAGGCAGC TTGAAAGGCG ATACCAACAT  
35 501 TGTTCAGGC AAACACTACG AACAAATCGG CAGTACCGTT TCCAGCCCGG  
551 AAGGCAACAA TACCATCTAT GCCCAAAGCA TAGACATTCA AGCGGCACAC  
601 AACAAATTAA ACAGTAATAC CACCCAAACC TATGAACAAA AAG. CTAAC  
651 GGTGCATTC AGTTCGCCCC TTACCGATTT GGCACAACAA ...

This corresponds to the amino acid sequence <SEQ ID 518; ORF117>:

40 1 ..SGNNLNAKAA EVSSANGTLA VSANNDINIS AGINTTHVDD ASKHTGRSGG  
51 GNKLVIDKA QSHHETAQSS TFEGKQVVVLQ AGNDANILGS NVISDNGTQI  
101 QAQGNHVRIGT TQTQSQSEY HQTQKSGLMS AGIGFTIGSK TNTQENQSQS  
151 NEHTGSTVGS LKGDTTIVAG KHYEQIGSTV SPEGNNITY AQSIDIAAH  
201 NKLNSNTTQT YEQKXLTVAF SSPVTDLAQQ ...

45 Computer analysis of this amino acid sequence gave the following results:

Homology with the pspA putative secreted protein of *N.meningitidis* (accession number AF030941)

ORF117 and pspA protein show 45% aa identity in 224aa overlap:

50 Orf117: 4 NLNAKAAEVSSANGTLAVSANNDINISAGINTTHVDDASKHTGRSGGGNKLVIDKAQSH 63  
++ +AAEV S G L ++A DI + AG T +DA K+TGRSGGG K +T ++  
pspA: 1173 DIRIRAAEVGSEQRLKLAAGRDIKVEAGKAHTETEDALKYTGRSGGGIKQKMRHLKNQ 1232

Orf117: 64 HETAQSSTFEGKQVVVLQAGNDANILGSNVISDNGTQIQAGNHVRIGTTQTQSQSEYHQT 123  
+ A S T +GK+++L +G D + GSN+I+DN T + A N++ + +T+S+S ++  
pspA: 1233 NGQAVSGTLDGKEIILVSGRDITVTGNSNIIADNHTILSAKNNIVLKAETR SRSAEMNKK 1292

55





601 AATGGCGCGA CTGCGGCAGG TTCGATGAAT CTCAGCGTTG GCATTGCATT  
 651 AAGTGCCGAG CAAGCAGCGC AACTGACCAG CGATATTGTT TGGTTGTGAC  
 701 AAAAAGAAGT TAAACTTCCT GATGGCGGCA CACAAACCGT ATTGATGCCA  
 751 CAGGTTTATG TACGCGTTAA AAATGGCGGC ATAGACGGTA AAGGTGCATT  
 801 GTTGTCAAGC AGCAATACAC AAATCAATGT TTCAGGCAGC CTGAAAAACT  
 851 CAGGCACGAT TGCAGGGCGC AATGCGCTTA TTATCAATAC CGATACGCTA  
 901 GACAATATCG GTGGGCGTAT TCATGCGCAA AAATCAGCGG TTACGGCCAC  
 951 ACAAGACATC AATAATATTG GCGGCATTCT TTCTGCCGAA CAGACATTAT  
 1001 TGCTCAATGC GGGTAACAAC ATCAACAACC AAAGCACGGC CAAGAGCAGT  
 1051 CAAAATGCAC AAGGTAGCAG CACCTACCTA GACCGAATGG CAGGTATTTA  
 1101 TATCACAGGC AAAGAAAAAG GTGTTTTAGC AGCGCAGGCA GGCAAAGACA  
 1151 TCAACATCAT TGCCGGTCAA ATCAGCAATC AATCAGATCA AGGGCAAACC  
 1201 CGGCTGCAGG CAGGACGCGA CATTAACCTG GATACGGTAC AAACCGGCAA  
 1251 ATATCAAGAA ATCCATTGTT ATGCCGATAA CCATACCATC CGAGGTTCAA  
 1301 CGAACGAAGT CGGCAGCAGC ATTCAAACAA AAGGCGATGT TACCCtatTG  
 1351 TCAGGGAATA ATCTCAATGC CAAAGCTGCC GAAGTCGGCA GCGCAAAGG  
 1401 CACACTTGCC GTGTATGCTA AAAATGACAT TACTATCAGC TCAGGCATCC  
 1451 ATGCCGCGCA AGTTGATGAT GCGTCCAAAC ATACAGGCAG AAGCGGCGGC  
 1501 GGTAATAAAT TAGTCATTAC CGATAAAGCC CAAAGTCATC ACGAAATCTG  
 1551 TCAAAGCAGC ACCTTTGAAG GCAAGCAAGT TGTATTGCAG GCAGGAAACG  
 1601 ATGCCAACAT CCTTGGCAGT AATGTTATTT CCGATAATGG CACCCGGATT  
 1651 CAAGCAGGCA ATCATGTTTCG CATTGGTACA ACCCAAATC AAAGCCAAAG  
 1701 CGAAACCTAT CATCAAACCC AAAAATCAGG ATTGATGAGT GCAGGTATCG  
 1751 GCTTCACTAT TGGCAGCAAG ACAAACACAC AAGAAAACCA ATCCCAAAGC  
 1801 AACGAACATA CAGGCAGTAC CGTAGGCAGC CTGAAAGGCG ATACCACCAT  
 1851 TGTGCAAGC AAACACTACG AACAAACCGG CAGCAACGTT TCCAGCCCTG  
 1901 AGGGCAACAA CCTTATCAGC ACGCAAAGTA TGGATATTGG CGCAGCACAA  
 1951 AACCAATTAA ACAGCAAAAC CACCCAAACC TACGAACAAA AAGGCTTAAC  
 2001 GGTGGCATTC AGTTCGCCCG TTACCGATTT GGCACAACAA GCGATTGCCG  
 2051 TAGCACACAA AGCAGCAAAAC AAGTCGGACA AAGCAAAAAC GACCGCGTTA  
 2101 ATGCCATGGC GGCTGCCAAT GCAGGTTGGC AGGCCTATCA AACAGGCAAA  
 2151 GGCGCACAAA ACTTAG

This corresponds to the amino acid sequence <SEQ ID 522; ORF117ng-1>:

1 LLVQTEKDGL HNEQTFGEKK VFSENGKLHN YWRARRKGHD ETGHREQNYT  
 51 LPPEITRDIS LGSFAYESH KALSRHAPSQ GTELPQSNRD NIRTAKSNGI  
 101 SLPYTPNSFT PLPGSSLYII NPANKGYLVE TDPFRFANYRQ WLGS DYMLGS  
 151 LKLDPNLHK RLGDGYEQR LINEQIAELT GHRRLDGYQN DEEQFKALMD  
 201 NGATAARSMN LSVGIALSAE QAAQLTSDIV WLWQKEVKLP DGGTQTVLMP  
 251 QYVVRVKNGG IDKGALLSG SNTQINVSGS LKNSGTIAGR NALIINTDTL  
 301 DNIGGRIHAQ KSAVTATQDI NNIGGILSAE QTLNLLNAGNN INNQSTAKSS  
 351 QNAQGSSTYL DRMAGIYITG KEGVLAAQA GKDINI IAGQ ISNQSDQQT  
 401 RLQAGRDINL DTVTGKYQE IHFDADNHTI RGSTNEVGSS IQTKGDVTL  
 451 SGNLNAKAA EVGSAKGLA VYAKNDITIS SGIHAGQVDD ASKHTGRSGG  
 501 GNKLVIDDKA QSHHETAQSS TFEKGQVVLQ AGNDANILGS NVISDNGTRI  
 551 QAGNHVRIGT TQTQSQSEY HQTQKSLMS AGIGFTIGSK TNTQENQSQS  
 601 NEHTGSTVGS LKGDFTIVAS KHYEQTGSNV SPEGNNLIS TQSM DIGAAQ  
 651 NQLNSKTTQT YEQKGLTVAF SSPVTDLAQQ AIAVAHKAAN KSDKAKTTAL  
 701 MPWRLPMQVG RPIKQAKAHK T\*

ORF117ng-1 shows the same 90% identity over a 230aa overlap with ORF117. In addition, it shows homology with a secreted *N. meningitidis* protein in the database:

gi|2623258 (AF030941) putative secreted protein [Neisseria meningitidis] Length = 2273

Score = 604 bits (1541), Expect = e-172

Identities = 325/678 (47%), Positives = 449/678 (65%), Gaps = 22/678 (3%)

Query: 1 LLVQTEKDGLHNEQTFGEKKVFSENGKLHNYWRARRKGHD ETGHREQNYTLPEEITRDIS 60  
 L+V T + L N++T G K + ++ G L H Y R +KG D TG+ Y E++ I  
 Sbjct: 739 LIVGTPESALDND ETLGKTI-TDKGDLHRYHRHKKGRDSTGYSRSPYEPAPEVS-SIR 796

Query: 61 LGSFAYESHKALSRHAPSQGT ELPQSNRDNIRTAKSNGISLPYTPNSFTPLPGSSLYII 120  
 +G AY+ + AP Q +++P + + NGI +T LP SSL+ I  
 Sbjct: 797 MGISAYKGY-----APQASDIPGTV---VPVVAENG IHTFT-----LPNSSLFAI 840

Query: 121 NPANKGYLVETDPRFANYRQWLGS DYMLGSLKLDPNLHKRLGDGYEQR LINEQIAELT 180

Sbjct: 841 P NKGYL+ETDP F +YR+WLGS YML +L+ DPN++HKRLGDGYEQL+NEQIA+LT  
APNNKGYLIETDPAFTDYRKWLGSYMLAALQQDPNHIHKRLGDGYEQLVNEQIAKLT 900

Query: 181 GHRRLDGYQNDEEQFKALMDNGATAARSMNLSVGIALSAEQAAQLTSDIVWLQKEVKLP 240  
G+RRLDGY NDEEQFKALMDNG T A+ + L+ GIALSAEQ A+LTSDIVWL + V LP

Sbjct: 901 GYRRLDGYTNDEEQFKALMDNGITIAKELQLTPGIALSAEQVARLTSDIVWLENETVTLT 960

Query: 241 DGGTQTVLMPQVYVRVKNGGIDGKGALLSGSNTQINVSGSLKN-SGTIAGRNALIINTDT 299  
DG TQTVL P+VYVR + ++G+GALLSGS I SG+++N G IAGR ALI+N

Sbjct: 961 DGTQTVLKPQVYVRARPKDMNGQGALLSGSVVDIG-SGAIENRGGLIAGREALILNAQN 1019

Query: 300 LDNIGGRIHAQKSAVTATQDINNIGGILSAEQTLNAGNNINNOSTAKSSQNAQGSSTY 359  
+ N+ G + + A DI N G I AE LLL A NNI ++S +S+QN QGS

Sbjct: 1020 IKNLQGDLDQKNIFAAAGSDITNTGSI-GAENALLKASNNIESRSETRSNQNEQGSVRN 1078

Query: 360 LDRMAGIYITGKEKGVLAQAQAGKDINIIAGQISNQSDQGTQLQAGRDNLDTVQTKGYQ 419  
+ R+AGIY+TG++ G + AG +I + A +++NQ+ GQT L AG DI DT + Q

Sbjct: 1079 IGRVAGIYLTGRQNGSVLLDAGNNIVLTASELTNQSEDGQTVLNAGGDIRSDTTGISRNQ 1138

Query: 420 EIHFDADNHTIRGSTNEVGSSIQTKGDVTLNAGNNLNKAAAEVGSAGKTLAVYAKNDITI 479  
FD+DN+ IR NEVGS+I+T+G+++L + ++ +AAEVGS +G L + A DI +

Sbjct: 1139 NTIFDSNDYVIRKEQNEVGSTIRTRGNLSLNAGKDIRIRAAEVGSEQRLKLAAGRDIKV 1198

Query: 480 SSGIHAGQVDDASKHTGRSGGKLVITDKAQSHHETAQSSTFEGKQVVLQAGNDANILG 539  
+G + +DA K+TGRSGGG K +T ++ + A S T +GK+++L +G D + G

Sbjct: 1199 EAGKAHTETEDALKYTGRSGGGIKQKMRHLKNQNGQAVSGTLDGKEIILVSGRDITVTG 1258

Query: 540 SNVISDNTRIAGNHVRIGTTQTSQSETYHQTQKSGLM-SAGIGFTIGSKTNTQENQS 598  
SN+I+DN T + A N++ + +T+S+S ++ +KSGLM S GIGFT GSK +TQ N+S

Sbjct: 1259 SNIIADNHTILSAKNNIVLKAETR SRSAEMNKKEKSGLMGSGGIGFTAGSKKDTQTNR 1318

Query: 599 QSNEHTGSTVSGSLKGDTTIVASKHYEQTSNVSSPEGNNLISTQSMDIGAAQNQLNSKTT 658  
++ HT S VGSL G+T I A KHY QTGS +SSP+G+ IS+ + I AQN+ + ++

Sbjct: 1319 ETVSHTESVVGSLNGNTLISAGKHYTQTGSTISSPQGDVGISSGKISIDAAQNRYSQESK 1378

Query: 659 QTYEQKGLTVAFFSSPVT 676  
Q YEQKG+TVA S PV +

Sbjct: 1379 QVYEQKGVTVVAISVPV 1396

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 63

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 523>:

1 ATGATTTACA TCGTACTGTT TCTAGCTGTC GTCCTCGCCG TTGTGCCTA  
51 CAACATGTAT CAGGAAAACC AATACCGCAA AAAAGTGCGC GACCAGTTCG  
101 GACACTCCGA CAAAGATGCC CTGCTCAACA GCAwAACAG CCATGTCCGC  
151 GACGGCAAAC CGTCCGGCGG GTCAGTCATG ATGCCGAAAC CCAACCGGC  
201 GGTCAAAAAA ACGGCAAAC CCAAGACCC CGyCATGCGC AACCTGCAAG  
251 AACAGGATGC CGTCTACATC GCCAAGCAGA AACAGGCAA AGCCTCCCCG  
301 TTCAAAACCG AAATCGAAAC CGCCTGGAA GAAAGCGCA TTATCGGCA  
351 CTCCGCCCAC ACCGTTTCCG AACCCCAAAC CGGACATTCC GCAACGAAAC  
401 CTGCCGACGC GTCGGCAAAC CCTGCACCCG TTCCGCAAAC ACCTGCAAAA  
451 CCGCTGATTA CGCTCAAAGA ACTGTCAAAA GTCGAATTAT CCTGGTTTGA  
501 CGTGCGCATC GACTTCATCT CCTAT...

This corresponds to the amino acid sequence <SEQ ID 524; ORF119>:

1 MIYIVLFLAV VLAVVAYNMY QENQYRKKVR DQFGHSDKDA LLNSXTSHVR  
51 DGKPSGGSVM MPKPQPAVKK TAKPDPMXR NLQEQDAVYI AKQKQAKASP  
101 FKTEIETALE ESGIIGNSAH TVSEPQTGHS ATKPADASAK PAPVPQTPAK  
151 PLITLKELSK VELSWFDVRI DFISY...

Further work revealed the complete nucleotide sequence <SEQ ID 525>:

1 ATGATTTACA TCGTACTGTT TCTAGCTGTC GTCCTCGCCG TTGTCGCCTA  
 51 CAACATGTAT CAGGAAAACC AATACCGCAA AAAAGTGGCG GACCAGTTCCG  
 101 GACACTCCGA CAAAGATGCC CTGCTCAACA GCAAAACCAG CCATGTCCGC  
 151 GACGGCAAAC CGTCCGGCGG GTCAGTCATG ATGCCGAAAC CCCAACCGGC  
 5 GGTCAAAAAA ACGGCAAAAC CCCAAGACCC CGCCATGCGC AACCTGCAAG  
 251 AACAGGATGC CGTCTACATC GCCAAGCAGA AACAGGCAA AGCCTCCCCG  
 301 TTCAAACCG AAATCGAAAC CGCCTTGGA GAAAGCGGCA TTATCGGCAA  
 351 CTCCGCCCAC ACCGTTTCCG AACCCCAAAC CGGACATTCC GCACCGAAAC  
 401 CTGCGGACGC GCCGGCAAAA CCTGCACCCG TTCGCAAAAC ACCTGCAAAA  
 10 451 CCGCTGATTA CGCTCAAAGA ACTGTCAAAA GTCGAATTAC CCTGGTTTGA  
 501 CGTGCCTTTC GACTTCATCT CCTATATCGC GCTGACCGAA GCCAAGAAGC  
 551 TGCACGCACT GCCGCGCCTT TCCAACCGCT GCCGCTACCA GATTGTCCGC  
 601 TGCACCATGG ACGACCATTT CCAGATTGCC GAACCCATCC CGGGCATCCG  
 651 CTATCAGGCA TTTATCGTGG GTATTCAAGC AGTCAGCCGC AACGGATTG  
 15 701 CCGTGCAGGA AGAACTCTCC GCATTCAACC GCCAGGTGGA CGCATTCGCA  
 751 CAAAGCATGG GCGGTGAGC GCTGCACACC GACCTTGCCG CCTTTATCGA  
 801 AGTGGCTTCC GCACTGGACG CATTTCTGCG GCGCGTCGAC CAGACCATCG  
 851 CCATCCATTT GGTTTCCCG ACCAGCATCA GCGGCGTAGA ACTGCGTTCC  
 901 GCCGTAACGG GCGTGGGTTT CGTTTTGGAA GACGACGGCG CGTTCCACTA  
 20 951 TACCGACACG TCGGGCTCGA CCATGTTCTC CATCTGCTCG CTCAACAACG  
 1001 AGCCGTTTAC CAACGCCCTT TTGGACAACC AGTCCTACAA AGGCTTCAGT  
 1051 ATGCTGCTCG ACATCCGCA CTCTCCGCA GCGCAAAAAA CCTTCGACGA  
 1101 TTTGTTTATG GATTGGCGG TACGCCTGTC CGGCCAGTTG AACCTGAATC  
 1151 TGGTCAACGA CAAAATGGAA GAAGTTTCGA CCAATGGCT CAAAGACGTG  
 25 1201 CGCACTTATG TATTGGCGCG TCAGTCCGAG ATGCTCAAAG TCGGTATCGA  
 1251 ACCGGGCGGC AAAACCGCAT TCGCCTGTT CTCCTAA

This corresponds to the amino acid sequence <SEQ ID 526; ORF119-1>:

1 MIYIVLFLAV VLAVVAYNMY QENQYRKKVR DQFGHSDKDA LLNSKTSHVR  
 51 DGKPSGGGSM MPKPQPAVKK TAKPQDPAMR NLQEQDAVYI AKQKQAKASP  
 101 FKTEIETALE ESGIIGNSAH TVSEPQTGHS APKPADAPAK PAPVPQTPAK  
 151 PLITLKELSK VELPWFVRF DFISYIALTE AKELHALPRL SNRCRYQIVG  
 201 CTMDDHFQIA EPIPGIRYQA FIVGIQAVSR NGLASQEELS AFNRQVDAFA  
 251 QSMGGQTLHT DLAAFIEVAS ALDAFCARVD QTIAIHLVSP TSISGYVELRS  
 301 AVTGVGVLE DDGAFHYTDT SGTSMFISCS LNNPFTNAL LDNQSYKGFS  
 35 351 MLLDIPHSPA GEKTFDDLFM DLAVRLSGQL NLNLVNDKME EVSTQWLKDV  
 401 RTYVILARQSE MLKVGIEPPG KTAIRLFS\*

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF119 shows 93.7% identity over a 175aa overlap with an ORF (ORF119a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
orf119.pep		MIYIVLFLAVVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSXTSHVRDGKPSGGGSM					
orf119a		:					
		10	20	30	40	50	60
orf119.pep		MPKPQPAVKKTAKPQDPXMRNLQEQDAVYIAKQKQAKASPFKTEIETALES					
orf119a							
		70	80	90	100	110	120
orf119.pep		MPKPQPAVKKTAKPQDPXMRNLQEQDAVYIAKQKQAKASPFKTEIETALES					
orf119a							
		70	80	90	100	110	120
orf119.pep		TVSEPQTGHSATKPADASAKPAPVPQTPAKPLITLKELSKVELSWDFVRI					
orf119a							
		130	140	150	160	170	
orf119a		TVPEPQTGHSAPKPADAPAKPVPVPQTPAKPLITLKELSKVELPWFVRF					
		130	140	150	160	170	180
orf119a		AKELHALPRLSNRCRYQIVGCTMDDHFQIAEPIPGIRYQAFIVGIQAVSRNGLASQEELS					
		190	200	210	220	230	240

The complete length ORF119a nucleotide sequence <SEQ ID 527> is:

1 ATGATTTACA TCGTACTGTT CCTCGCCGCC GTCCTCGCCG TTGTCGCCTA  
51 CAATATGTAT CAGGAAAACC AATACCGCAA AAAAGTGC GC GACCAGTTTCG  
101 GGCACCTCCGA CAAAGATGCC CTGCTCAACA GCAAAACCAG CCATGTCCGC  
151 GACGGCAAAC CGTCCGGCGG GCCAGTCATG ATGCCGAAAC CCCAACCGGC  
5 GGTCAAAAAA ACGGCAAAAT CCCAAGACCC CGCCATGCGC AACCTGCAAG  
251 AGCAGGATGC CGTCTACATC GCCAAGCAGA AACAGGCAAA AGCCTCCCCG  
301 TTCAAAACCG AAATCGAAAC CGCCTTGGA GAAAGCGCA TTATCGGCAA  
351 CTCCGCCAC ACCGTCCCCG AACCCCAAAC CGGACATTCC GCACCAAAAC  
401 CTGCGGACGC GCCGGCAAAA CCTGTTCCCG TTCCGCAAAAC GCCGGCAAAA  
10 CCGCTGATTA CGCTCAAAGA GCTGTCGAAG GTCGAGCTGC CCTGGTTTGA  
451 CGTGCCTTC GACTTCATCT CTTATATCGC GCTGACCGAA GCCAAAGAAC  
501 TGCACGCACT GCCGCGCCTT TCCAACCGCT GCCGCTACCA GATTGTCGGC  
551 TGCACCATGG ACGACCATTT CCAGATTGCC GAACCCATCC CGGGCATCCG  
601 CTATCAGGCA TTTATCGTGG GTATTCAAGC AGTCAGCGC AACGGACTTG  
651 CCTCGCAGGA AGAATCTCC GCATTCAACC GCCAGGTGGA TGCATTGCGA  
15 CACAGCATGG CCGGTCAGAC GCTGCACACC GACCTTGCCG CCTTTATCGA  
751 AGTGGCTTCC GCACTGGACG CATTCTGCGC GCGCGTCGAC CAGACTATCG  
801 CCATCCATTT GGTTCCTCCG ACCAGCATCA GCGGCGTAGA ACTGCGTTCC  
851 GCCGTAACGG GCGTGGGTTT CGTTTTGGAA GACGACGGCG CGTTCCACTA  
901 TACCGACACG TCGGGCTCGA CCATGTTCTC CATCTGCTCG CTCACAACG  
20 AGCCGTTTAC CAATGCCCTT TTGGACAACC AGTCCTATAA AGGCTTCAGT  
1001 ATGCTGCTCG ACATCCCGCA CTCTCCGCA GCGGAAAAA CCTTCGACGA  
1051 TTTGTTTATG GATTGCGCG TACGCCTGTC CGGCCAGTTG AACCTGAATC  
1101 TGGTCAACGA CAAAATGGAA GAAGTTTCGA CCAATGGCT CAAAGACGTG  
1151 CGCACTTATG TATTGGCTCG TCAGTCCGAG ATGCTCAAAG TCGGTATCGA  
25 ACCGGCGCG AAAACCGCAT TCGCCTGTT CTCTAA

This encodes a protein having amino acid sequence <SEQ ID 528>:

1 MIYIVLFLAA VLAVVAYNMY QENQYRKKVR DQFGHSDKDA LLNSKTSVHR  
51 DGKPSGGPVM MPKPQPAVKK TAKSQDPAMR NLQEQDAVYI AKQKQAKASP  
30 FKTEIETALE ESGIIGNSAH TVPEPQTGHS APKPADAPAK PVPVPQTPAK  
151 PLITLKELSK VELPWFDFVRF DFISYIALTE AKELHALPRL SNRCRYQIVG  
201 CTMDDHFQIA EPIPGIRYQA FIVGIQAVSR NGLASQEELS AFNRQVDAFA  
251 HSMGGQTLHT DLAFFIEVAS ALDAFCARVD QTIAIHLVSP TSISGVELRS  
301 AVTGVGFVLE DDGAFHYTDT SGSTMFSICS LNNPFTNAL LDNQSYKGFV  
351 MLLDIPHSPA GEKTFDDLEM DLAVRLSGQL NLNLVNDKME EVSTQWLKDV  
401 RTYVLRQSE MLKVGIEPGG KTALRLFS\*

ORF119a and ORF119-1 show 98.6% identity in 428 aa overlap:

		10	20	30	40	50	60	
40	orf119a.pep	MIYIVLFLAAVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSVHRD	DGKPSGGPVM					
	orf119-1	MIYIVLFLAVVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSVHRD	DGKPSGGPVM					
		10	20	30	40	50	60	
45	orf119a.pep	MPKPQPAVKKTAKSQDPAMRN	70	80	90	100	110	120
	orf119-1	MPKPQPAVKKTAKSQDPAMRN	70	80	90	100	110	120
		130	140	150	160	170	180	
50	orf119a.pep	TVPEPQTGHSAPKPADAPAKPVPV	QTPAKPLITLKELSKVELPWFDFVRFDFISYIALTE					
	orf119-1	TVSEPQTGHSAPKPADAPAKPAPVP	QTPAKPLITLKELSKVELPWFDFVRFDFISYIALTE					
55		130	140	150	160	170	180	
	orf119a.pep	AKELHALPRLSNRCRYQIVGCTMDDHFQIAEPIPGIRYQAFIVGIQAVSRNGLASQEELS						
	orf119-1	AKELHALPRLSNRCRYQIVGCTMDDHFQIAEPIPGIRYQAFIVGIQAVSRNGLASQEELS						
60		190	200	210	220	230	240	
	orf119a.pep	AFNRQVDAFAHSMGGQTLHTDLA	AFIEVASALDAFCARVDQTIAIHLVSP					
65	orf119-1	AFNRQVDAFAHSMGGQTLHTDLA	AFIEVASALDAFCARVDQTIAIHLVSP					
		250	260	270	280	290	300	

		310	320	330	340	350	360
	orf119a.pep	AVTGVGFVLEDDGAFHYTDTSGSTMFSICSLNNEPFTNALLDNQSYKGFMSMLLDIPHSPA					
5	orf119-1	AVTGVGFVLEDDGAFHYTDTSGSTMFSICSLNNEPFTNALLDNQSYKGFMSMLLDIPHSPA					
		310	320	330	340	350	360
	orf119a.pep	GEKTFDDLFDLAVRLSGQLNLNLVNDKMEEVSTQWLKDVRTYVVLARQSEMLKVGIEP					
10	orf119-1	GEKTFDDLFDLAVRLSGQLNLNLVNDKMEEVSTQWLKDVRTYVVLARQSEMLKVGIEP					
		370	380	390	400	410	420
	orf119a.pep	KTALRLFSX					
15	orf119-1	KTALRLFSX					

### Homology with a predicted ORF from *N.gonorrhoeae*

- 20 ORF119 shows 93.1% identity over a 175aa overlap with a predicted ORF (ORF119ng) from *N.gonorrhoeae*:

	orf119.pep	MIYIVLFLAVVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSXTSHVRD	60
	orf119ng	MIYIVLFLAAVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSHVRD	60
25	orf119.pep	MPKPQPAVKKTA	120
	orf119ng	MPKPQPAVKKPAKPQDSAMRN	120
30	orf119.pep	TVSEPQTGHSATKPADASAKPAPVPQTPAKPLITLKELSKVELSWFDVRIDFISY	175
	orf119ng	TVSEPQTGHSAPKPADAPAKPVPVPQTPAKPLITLKELSKVELPWFDVREDFISYIALTE	180

The complete length ORF119ng nucleotide sequence <SEQ ID 529> is:

35	1	ATGATTTACA	TCGTA	CTCGCCGCC	GTCCTCGCCG	TTGTG
	51	CAATATGTAT	CAGGAAAACC	AATACCGCAA	AAAAGTGCGC	GACCA
	101	GACACTCCGA	CAAAGATGCC	CTGCTCAACA	GCAAACCCAG	CCATG
	151	GACGGCAAAC	CGTCCGGCGG	GCCAGTCATG	ATGCCGAAAC	CCCAAC
	201	GGTCAAAAAA	CCGGCCAAAC	CCCAAGACTC	CGCCATGCGC	AACCTG
	251	AACAGGATGC	CGCTACATC	GCCAAGCAGA	AACAGGCAAA	AGCCTC
40	301	TTCAAAACCG	AAATCGAAAC	CGCCTTGGA	GAAATCGGCA	TTATCG
	351	CTCCGCCAC	ACCGTTTCCG	AACCCCAAAC	CGGACATTC	GCACCG
	401	CTGCCGACG	GCCGGCAAAA	CCGTTCCCG	TTCCGCAAAC	GCCGGC
	451	CCGCTGATTA	CGCTCAAAGA	GCTGTGGAAG	GTCGAGCTGC	CCTGGT
45	501	CGTGCGCTtc	gACTTCATCT	CCTATATCGC	GCTGACCGAA	GCCAAAG
	551	TGCACGCACT	GCCGCGCCTT	tcAACC	GCCGCTACCA	GATTGT
	601	TGCACCATGG	ACGACCATTT	CCAGATTGCC	GAACCCATCC	CGGGCAT
	651	CTATCAGGCA	TTATCGTGG	GTATCCAGGC	AGTCAGCCGC	AACGGACT
	701	CCTCGCAGGA	AGAACTCTCC	GCATTC	GCCAGGCGGA	CGCATTC
50	751	CAAAGCATGG	GCGGTCAGAC	GCTGCACACC	GACCTTGCCG	CCTTTAT
	801	AGTGGCTTCC	GCACTGGACG	CATTCTGCGC	GCGCGTCGAC	CAGACCAT
	851	CCATCCATTT	GGTTTCGCCG	ACCAGCATCA	GCGGCGTAGA	ACTGCGT
	901	GCCGTAACGG	GCGTGGGTTT	CGTTTGGAA	GACGACGGCG	CGTCCACTA
	951	TACCGACACG	TCGGGCTCGA	CCATGTTCTC	CATCTGCTCG	CTCAACA
55	1001	AGCCGTTTAC	CAATGCCCTT	TTGACAAC	AGTCCTACAA	AGGCTT
	1051	ATGCTGCTCG	ACATCCCGCA	CTCTCCGGCA	GGCGAAAAAA	CCTTCG
	1101	TTGTTTATG	GATTTGGCGG	TACGCTGTC	CGGTCAAGTTG	AACCTGA
	1151	TGGTCAACGA	CAAAATGGAA	GAAGTTTCGA	CCCAATGGCT	CAAAGAC
	1201	CGCACTTATG	TATTGGCGCG	TCAGTCCGAG	ATGCTCAAAG	TCGGTAT
	1251	ACCGGGCGGC	AAAACCGCCC	TGCGCCTGTT	TTCATAA	

- 60 This encodes a protein having amino acid sequence <SEQ ID 530>:

1	MIYIVLFLAA	VLAVVAYNMY	QENQYRKKVR	DQFGHSDKDA	LLNSKTSHVR
51	DGKPSGGPVM	MPKPQPAVKK	PAKPQDSAMR	NLQEQDAVYI	AKQKQAKASP

5  
 101 FKTEIETALE EIGIIGNSAH TVSEPQTGHS APKPADAPAK PVPVPQTPAK  
 151 PLITLKELSK VELPWFDVRF DFISYIALTE AKELHALPRL SNRCRYQIVG  
 201 CTMDHDFQIA EPIPGIRYQA FIVGIQAVSR NGLASQEELS AFNRQADAF  
 251 QSMGGQTLHT DLAAFIEVAS ALDAFCARVD QTIHILVSP TSISGVELRS  
 301 AVTGVGVFVLE DDGAFHYTDT SGSTMFSICS LNNEPFTNAL LDNQSYKGF  
 351 MLLDIPHSPA GEKTFDDLFM DLAVRLSGQL NLNLVNDKME EVSTQWLKDV  
 401 RTYVLARQSE MLKVGIEPPG KTALRLFS\*

ORF119ng and ORF119-1 show 98.4% identity over 428 aa overlap:

10  
 orf119ng                    10                    20                    30                    40                    50                    60  
 MIYIVLFLAAVLAVVAYNMYQENQYRKVVDQFGHSDKDALLNSKTSHVVDGKPSGGPVM  
 |||||:|||||  
 orf119-1                    10                    20                    30                    40                    50                    60  
 MIYIVLFLAVVLAVVAYNMYQENQYRKVVDQFGHSDKDALLNSKTSHVVDGKPSGGSV  
 15  
 orf119ng                    70                    80                    90                    100                    110                    120  
 MPKPQPAVKKPAKQDSAMRNLEQDAVYIAKQKQAKASPFFKTEIETALEEIGIIGNSAH  
 |||||:|||||  
 orf119-1                    70                    80                    90                    100                    110                    120  
 MPKPQPAVKKTAKPQDPAMRNLEQDAVYIAKQKQAKASPFFKTEIETALEESGIIGNSAH  
 20  
 orf119ng                    130                    140                    150                    160                    170                    180  
 TVSEPQTGHSAPKPADAPAKPVPVQTPAKPLITLKELSKVELPWFDVRFDFISYIALTE  
 |||||:|||||  
 orf119-1                    130                    140                    150                    160                    170                    180  
 TVSEPQTGHSAPKPADAPAKPVPVQTPAKPLITLKELSKVELPWFDVRFDFISYIALTE  
 25  
 orf119ng                    190                    200                    210                    220                    230                    240  
 AKELHALPRLSNRCRYQIVGCTMDHDFQIAEPIPGIRYQAFIVGIQAVSRNGLASQEELS  
 |||||:|||||  
 orf119-1                    190                    200                    210                    220                    230                    240  
 AKELHALPRLSNRCRYQIVGCTMDHDFQIAEPIPGIRYQAFIVGIQAVSRNGLASQEELS  
 30  
 orf119ng                    250                    260                    270                    280                    290                    300  
 AFNRQADAFQSMGGQTLHTDLAAFIEVASALDAFCARVDQTIHILVSPSTISGVELRS  
 |||||:|||||  
 orf119-1                    250                    260                    270                    280                    290                    300  
 AFNRQVDAFAQSMGGQTLHTDLAAFIEVASALDAFCARVDQTIHILVSPSTISGVELRS  
 35  
 orf119ng                    310                    320                    330                    340                    350                    360  
 AVTGVGVFVLEDDGAFHYTDTSGSTMFSICSLNNEPFTNALLDNQSYKGFMSMLLDIPHSPA  
 |||||:|||||  
 orf119-1                    310                    320                    330                    340                    350                    360  
 AVTGVGVFVLEDDGAFHYTDTSGSTMFSICSLNNEPFTNALLDNQSYKGFMSMLLDIPHSPA  
 40  
 orf119ng                    370                    380                    390                    400                    410                    420  
 GEKTFDDLFDLAVRLSGQLNLNLVNDKMEEVSTQWLKDVRTYVLARQSEMLKVGIEPPG  
 |||||:|||||  
 orf119-1                    370                    380                    390                    400                    410                    420  
 GEKTFDDLFDLAVRLSGQLNLNLVNDKMEEVSTQWLKDVRTYVLARQSEMLKVGIEPPG  
 45  
 orf119ng                    429  
 KTALRLFSX  
 |||||  
 orf119-1                    429  
 KTALRLFSX

55 Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 64

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 531>

5  
10  
15  
20  
25  
30  
35  
40  
45

```

1  . . GCGCGGCACG GCACGGAAGA TTTCTTCATG AACACAGCG ACAC.ATCAG
51 GCAGATAGTC GAAAGCACCA CCGGTACGAT GAAGCTGCTG ATTTCTCTCA
101 TCGCCCTGAT TTCATTGGTA GTCGGCGGCA TCGGCGTGAT GAACATCATG
151 CTGGTGTCCG TTACCGAGCG CACCAAAGAA ATCGGCATAC GGATGGCAAT
201 CGGCGCGCGG CGCGGCAATA TTTyGCAGCA GTTTTGTGAT GAGGCGGTGT
251 TAATCTGCGT CATCGGCGGT TTGGTCGGCG TGGGTTTGTC CGCCGCCGTC
301 AGCCTCGTGT TCAATCATTT TGTAACCGAC TTCCCGATGG ACATTTCGCG
351 CATGTCCGTC ATCGGCGCGG TCGCCTGTTT GACCGGAATC GGCATCGCGT
401 TCGGCTTTAT GCCTGCCAAT AAAGCAGCCA AACTCAATCC GATAGACGCA
451 TTGGCACAGG ATTGA

```

This corresponds to the amino acid sequence <SEQ ID 532; ORF134>:

15  
20  
25  
30  
35  
40  
45

```

1  . . ARHGTEDFFM NNSDXIRQIV ESTTGMTKLL ISSIALISLV VGGIGVMNIM
51 LVSVTERTKE IGIRMAIGAR RGNIXQQFLI EAVLICVIGG LVGVGLSAAV
101 SLVFNHFVTD FPMDISAMSV IGAVACSTGI GIAFGFMPAN KAAKLNPIDA
151 LAQD*

```

Further work revealed the complete nucleotide sequence <SEQ ID 533>:

20  
25  
30  
35  
40  
45

```

1  ATGTCGGTGC AAGCAGTATT GGCGCACAAA ATGCGTTTCG TCTGACGAT
51 GGTTCGGCATC ATCATCGGTA TCGCGTCGGT GGTTCCTGTC GTCGCATTGG
101 GCAATGGTTC GCAGAAAAAA ATCCTTGAAG ACATCAGTTC GATAGGACG
151 AACACCATCA GCATCTTCCC GGGGCGCGGC TTCGGCGACA GGCGCAGCGG
201 CAGGATTAAA ACCTGACCA TAGACGACGC AAAAATCATC GCCAAACAAA
251 GCTACGTGTC TTCCGCCACG CCCATGACTT CGAGCGGCGG CACGCTGACT
301 TACCGCAACA CCGACCTGAC CGCCTCGCTT TACGGCGTGG GCGAACATA
351 TTTGACGTG CGCGGACTGA AGCTGGAAC GGGGCGGCTG TTTGACGAAA
401 ACGATGTGAA AGAAGACGCG CAGGTCGTCG TCATCGACCA AAATGTCAAA
451 GACAACTCTT TTGCGGACTC GGATCCGTTG GGTAAACCA TTTGTTGAG
501 GAAACGCCCC TTGACCGTCA TCGGCGTGAT GAAAAAGAC GAAACGCTT
551 TCGGCAATTC CGACGTGCTG ATGCTTTGGT CGCCCTATAC GACGGTGATG
601 CACCAATCA CAGGCGAGAG CCACACCAAC TCCATCACCG TCAAAATCAA
651 AGACAATGCC AATACCCAGG TTGCCGAAAA AGGGCTGACC GATCTGCTCA
701 AAGCGCGGCA CGGCACGGAA GATTTCTTCA TGAACAACAG CGACAGCATC
751 AGGCAGATAG TCGAAAGCAC CACCGGTACG ATGAAGCTGC TGATTTCTCTC
801 CATCGCCCTG ATTTATTGG TAGTCGGCGG CATCGGCGTG ATGAACATCA
851 TGCTGGTGTC CGTTACCGAG CGCACCAAG AAATCGGCAT ACGGATGCA
901 ATCGGCGCGC GCGCGGGCAA TATTTGTCAG CAGTTTTTGA TTGAGGCGGT
951 GTTAATCTGC GTATCGGCGG GTTTGGTTCG CGTGGGTTTG TCCGCGCGCG
1001 TCAGCCTCGT GTTCAATCAT TTTGTAACCG ACTTCCCGAT GGACATTTC
1051 GCCATGTCCG TCATCGGCGC GGTGCGCTGT TCGACCGGAA TCGGCATCGC
1101 GTTCGGCTTT ATGCTGCCA ATAAAGCAGC CAAACTCAAT CCGATAGACG
1151 CATTGGCACA GGATTGA

```

This corresponds to the amino acid sequence <SEQ ID 534; ORF134-1>:

45  
50  
55

```

1  MSVQAVLAHK MRSLLTMLGI IIGIASVSV VALNGSQKK ILEDISSIGT
51 NTISIFPGRG FGDRRSRIK TLTIDDAKII AKQSYVASAT PMTSSGGTLT
101 YRNTDLTASL YGVGEQYFDV RGLKLETGRL FDENDVKEDA QVVVIDQNVK
151 DKLFADSDPL GKTLFRKRP LTVIGVMKKD ENAFGNSDVL MLWSPYTTVM
201 HQITGESHTN SITVKIKDNA NTQVAEKGLT DLLKARHGTE DFFMNSDSI
251 RQIVESTTGT MKLLISSIAL ISLVVGGIGV MNIMLVSVTE RTKEIGIRMA
301 IGARRGNILQ QFLIEAVLIC VIGGLVGVGL SAAVSLVENH FVTDFPMDIS
351 AMSVIGAVAC STGIGIAFGF MPANKAAKLN PIDALAQD*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with the hypothetical protein o648 of *E.coli* (accession number AE000189)

ORF134 and o648 protein show 45% aa identity in 153aa overlap:

55

```

Orf134: 2  RHGTEDFFMNSDXIRQIVESTTGMTKXXXXXXXXXXXXXVGGIGVMNIMLVSVTERTKEI 61
          RHG +DEF N D + + VE TT T++ VGGIGVMNIMLVSVTERT+EI
o648: 496 RHGKKDFFTWNMDGVLTVEKTTRTLQLFLTLVAVISLVGGIGVMNIMLVSVTERTREI 555

Orf134: 62  GIRMAIGARRGNIXQQFLIEAXXXXXXXXXXXXXXXXXXFNHFVTD FPMDISAMSVI 121
          GIRMA+GAR ++ QQFLIEA F+ + + S +++
o648: 556 GIRMAVGARASDVLQQFLIEAVLVCLVGGALGITLSLLIAFTLQLFLPGWEIGFSPALL 615

```

Orf134: 122 GAVACSTGIGIAFGFMPANKAAKLNPIDALAQD 154  
A CST GI FG++PA AA+L+P+DALA++  
o648: 616 LAFLCSTVTGILFGWLPARNAARLDPVDALARE 648

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF134 shows 98.7% identity over a 154aa overlap with an ORF (ORF134a) from strain A of *N. meningitidis*:

		10	20	30	
10	orf134.pep	<div>ARHGTE<del>DFFMNNSDXIRQIVESTTGTMKLL</del></div>			
		<div>                                       </div>			
	orf134a	<div>GESH<del>TNSITVKIKDNANTQVAEKLTDLLKARHGTEDFFMNNSDSIRQIVESTTGTMKLL</del></div>			
		210	220	230	
		240	250	260	
15		40	50	60	
		70	80	90	
	orf134.pep	<div>ISSIALISLVS<del>VGGIGVMNIMLVSVTERTK</del>EIGIRMAIGAARRGNIXQQFLIEAVLICVIGG</div>			
		<div>   </div>			
	orf134a	<div>ISSIALISLVS<del>VGGIGVMNIMLVSVTERTK</del>EIGIRMAIGAARRGNILQQFLIEAVLICVIGG</div>			
		270	280	290	
		300	310	320	
20		100	110	120	
		130	140	150	
	orf134.pep	<div>LVGVLGLSA<del>AVSLVFNFHVTFDFPMDISAMSVIGAVACSTGIGIAFGFMPANKAAKLNPIDA</del></div>			
		<div>   </div>			
	orf134a	<div>LVGVLGLSA<del>AVSLVFNFHVTFDFPMDISAMSVIGAVACSTGIGIAFGFMPANKAAKLNPIDA</del></div>			
		330	340	350	
		360	370	380	
25					
	orf134.pep	<div>LAQDX        </div>			
30					
	orf134a	<div>LAQDX</div>			

The complete length ORF134a nucleotide sequence <SEQ ID 535> is:

	1	ATGTCGGTGC	AAGCAGTATT	GGCGCACAAA	ATGCGTTTCG	TTCTGACGAT
	51	GCTCGGCATC	ATCATCGGTA	TCGCTTCGGT	TGTCTCCGTC	GTGCGATTGG
35	101	GCAACGGTTC	GCAGAAAAAA	ATCCTTGAAG	ACATCAGTTT	GATAGGGGACG
	151	AACACCATCA	GCATCTTCCC	AGGGCGCGCG	TTCCGGCGACA	GGCGCGACGGG
	201	CAGGATTAAA	ACCTTGACCA	TAGACGACGC	AAAAATCATC	GCCAAACAAA
	251	GCTACGTTGC	TTCCGCCACG	CCCATGACTT	CGAGCGGCGG	CACGCTGACT
	301	TACCGCAATA	CCGACCTGAC	CGCTTCTTTG	TACGCTGTGG	GCGAACCAATA
40	351	TTTCGACGTG	CGCGGGCTGA	AGCTGGAAC	GGGCGCGCTG	TTTGACGAAA
	401	ACGATGTGAA	AGAAGACGCG	CAGGTCTGTC	TCATCGACCA	AAATGTCAAA
	451	GACAAACTCT	TTGCGGACTC	GGATCCGTTG	GGTAAACCA	TTTTGTTCAG
	501	GAAACGCCCC	TTGACCGTCA	TCGCGGTGAT	GAAAAAAGAC	GAAAACCGTT
	551	TCGGCAATT	CGACGTGCTG	ATGCTTTGGT	CGCCTATATC	GACGGTGATG
45	601	CACCAATCA	CAGCGAGAG	CCACACCAAC	TCCATCACCG	TCAAAATCAA
	651	AGACAATGCC	AATACCCAGG	TTGCCGAAAA	AGGGCTGACC	GATCTGTATC
	701	AAGCGCGGCA	CGGCACGGAA	GATTTCTTCA	TGAACAACAG	CGACAGCATC
	751	AGGCAGATAG	TCGAAAGCAC	CACCGGTACG	ATGAAGCTGT	TGATTTTCTC
	801	CATCGCCCTG	ATTTTCATTG	TAGTCGGCGG	CATCGGCGTG	ATGAACATCA
50	851	TGCTGTGTG	CGTTACCAG	CGCACCAAAG	AAATCGGCAT	ACGGATGGCA
	901	ATCGCGCGCG	GGCGCGGCCA	TATTTTGCAG	CAGTTTGTGA	TTGAGGCGGT
	951	GTTAACTCTG	GTCATCGGCG	GTTTGGTCGG	CGTGGGTTTG	TCCGCCGCCG
	1001	TCAGCCTCGT	GTTCAATCAT	TTTGTAACCG	ACTTCCCGAT	GGACATTTCC
	1051	GCCATGTCCG	TCATCGGCGC	GGTCGCCTGT	TCGACCGGAA	TCGGCATTCG
55	1101	GTTTCGGCTT	ATGCCTGCCA	ATAAAGCAGC	CAAACCTAAT	CCGATAGATG
	1151	CATTGGCGCA	GGATTGA			

This encodes a protein having amino acid sequence <SEQ ID 536>:

	1	MSVQAVLAHK	MRSLLTMLGI	IIGIASVVS	VALNGSQKK	ILEDISSIGT
	51	NTISIFPGR	FGDRSGRIK	TLTIDDAKII	AKQSVVASAT	PMTSSGGTLT
60	101	YRNTDLTASL	YGVGEQYFDV	RGLKLETGR	FEDENDKEDA	QVVVIDQNVK
	151	DKLFADSDPL	GKTILFRKRP	LTIVIGVMKKD	ENAFGNSDVL	MLWSPYTTVM
	201	HQITGESHTN	SITVKIKDNA	NTQVAEKGLT	DLLKARHGET	DFFMNNSDSI
	251	RQIVESTTGT	MKLLISSIAL	ISLVVGGIGV	MNIMLVSVTE	RTKEIGRMA



301 IGARRGNILQ QFLIEAVLIC VIGGLVGVGL SAAVSLVFNH FVTDFPMDIS  
351 AMSVIGAVAC STGIGIAFGF MPANKAKLN PIDALAQD\*

ORF134a and ORF134-1 show 100.0% identity in 388 aa overlap:

5	orf134a.pep	MSVQAVLAHKMRSLLTMLGIIIGIASVVSVALGNGSQKKILEDISSIGTNTISIFPGRG
	orf134-1	MSVQAVLAHKMRSLLTMLGIIIGIASVVSVALGNGSQKKILEDISSIGTNTISIFPGRG
10	orf134a.pep	FGDRRSGRIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV
	orf134-1	FGDRRSGRIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV
15	orf134a.pep	RGLKLETGRLFDENDVKEDAQVVVIDQNVKDKLFADSDPLGKTILFRKRPLTVIGVMKKD
	orf134-1	RGLKLETGRLFDENDVKEDAQVVVIDQNVKDKLFADSDPLGKTILFRKRPLTVIGVMKKD
20	orf134a.pep	ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTQVAEKGLTDLLKARHGTE
	orf134-1	ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTQVAEKGLTDLLKARHGTE
25	orf134a.pep	DFFMNNSDSIRQIVESTTGTMKLLISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMA
	orf134-1	DFFMNNSDSIRQIVESTTGTMKLLISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMA
30	orf134a.pep	IGARRGNILQQFLIEAVLICVIGGLVGVGLSAAVSLVFNHFVTDFPMDISAMSVIGAVAC
	orf134-1	IGARRGNILQQFLIEAVLICVIGGLVGVGLSAAVSLVFNHFVTDFPMDISAMSVIGAVAC
	orf134a.pep	STGIGIAFGFMPANKAKLNPIDALAQDX
	orf134-1	STGIGIAFGFMPANKAKLNPIDALAQDX

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF134 shows 96.8% identity over a 154aa overlap with a predicted ORF (ORF134.ng) from *N. gonorrhoeae*:

35	orf134.pep	ARHGTEDFFMNNSDXIRQIVESTTGTMKLL	30
	orf134ng	GESHTNSITVKIKDNANTRVAEKLGLAELLKARHGTEDFFMNNSDSIRQMVESTTGTMKLL	264
40	orf134.pep	ISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNIXQQFLIEAVLICVIGG	90
	orf134ng	ISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNILQQFLIEAVLICIIGG	324
45	orf134.pep	LVGVGLSAAVSLVFNHFVTDFPMDISAMSVIGAVACSTGIGIAFGFMPANKAKLNPIDA	150
	orf134ng	LVGVGLSAAVSLVFNHFVTDFPMDISAASVIGAVACSTGIGIAFGFMPANKAKLNPIDA	384
	orf134.pep	LAQD 154	
	orf134ng	LAQD 388	

50 The complete length ORF134ng nucleotide sequence <SEQ ID 537> is:

1	ATGTCGGTGC	AAGCAGTATT	GGCGCACAAA	ATGCGTTCGC	TTCTGACCAT
51	GCTCGGCATC	ATCATCGGTA	TCGCTTCGGT	TGTCTCCGTC	GTCCGCGTGG
101	GCAACGGTTC	GCAGAAAAAA	ATCCTCGAAG	ACATCAGTTC	GATGGGGACG
151	AACACCATCA	GCATCTTCCC	CGGGCGCGGC	TTCCGCGACA	GGCGCAGCGG
201	CAAAATCAAA	ACCCTGACCA	TAGACGACGC	AAAAATCATC	GCCAAACAAA
251	GCTACGTGTC	CTCCGCCACG	CCCATGACTT	CGAGCGGCGG	CACGCTGACC
301	TACCGCAATA	CCGACCTGAC	CGCTTCTTTG	TACGGTGTGG	GCGAACAATA
351	TTTCGACGTG	CGCGGGCTGA	AGCTGGAAC	GGGCGGCTG	TTTGATGAGA
401	ACGATGTGAA	AGAAGACGCG	CAAGTCGTCG	TCATCGACCA	AAATGTCAAA
451	GACAACTCT	TTGCGGACTC	GGATCCGTG	GGTAAAACCA	TTTTGTTCAG
501	GAAACGCCCC	TTGACCGTCA	TCGGCGTGAT	GAAAAAAGAC	GAAAACGCTT
551	TCGGCAATTC	CGACGTGCTG	ATGCTTTGGT	CGCCTATAC	GACGGTGATG

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60

```

601 CACCAAATCA CAGGCGAGAG CCACACCAAC TCCATCACCG TCAAAATCAA
651 AGACAATGCC AATACCCGGG TTGCCGAAAA AGGGCTGGCC GAGCTGCTCA
701 AAGCACGGCA CGGCACGGAA GACTTCTTTA TGAACAACAG CGACAGCATC
751 AGGCAGATGG TCGAAAGCAC CACCGGTACG ATGAAGCTGC TGATTTCTCT
801 CATCGCCCTG ATTTTCATTG TAGTCGCGCG CATCGGTGTG ATGAACATTA
851 TGCTGGTGTG CGTTACCGAG CGCACCAAAG AAATCGGCAT ACGGATGGCA
901 ATCGGCGCGC GGCGCGGCAA TATTTTGCAG CAGTTTTTGA TTGAGGCGGT
951 GTTAATCTGC ATCATCGGAG GCTTGGTCCG CGTAGGTTTG TCCGCCGCCG
1001 TCAGCCTCGT GTTCAATCAT TTTGTAACCG ATTTCCCGAT GGACATTTCC
1051 GCGGCATCCG TTATCGGGGC GGTGCGCTGT TCGACCGGAA TCGGCATCGC
1101 GTTCGGCTTT ATGCCTGCCA ATAAGGCAGC CAAACTCAAT CCGATAGATG
1151 CATTGGCGCA GGATTGA

```

This encodes a protein having amino acid sequence <SEQ ID 538>:

15  
20  
25  
30  
35  
40  
45  
50  
55  
60

```

1 MSVQAVLAHK MRSLLTMLGI IIGIASVSV VALNGSQKK ILEDISSMGT
51 NTISIFPGRG FGDRRSGKIK TLTIDDAKII AKQSYVASAT PMTSSGGTLT
101 YRNTDLTASL YGVGEQYFDV RGLKLETGRL FDENDVKEDA QVVVIDQNVK
151 DKLFAADSDPL GKTILFRKRP LTVIGVMKKD ENAFGNSDVL MLWSPYTTVM
201 HQITGESHTN SITVKIKDNA NTRVAEKGLA ELLKARHGTE DFFMNNSDSI
251 RQMVESTTGT MKLLISSIAL ISLVVGGIGV MNIMLVSVTE RTKEIGIRMA
301 IGARRGNILQ QFLIEAVLIC IIGGLVGVGL SAAVSLVFNH FVTDFFPMDIS
351 AASVIGAVAC STGIGIAFGF MPANKAAKLN PIDALAQD*

```

ORF134ng and ORF134-1 show 97.9% identity in 388 aa overlap:

25  
30  
35  
40  
45  
50  
55  
60

```

orf134ng MSVQAVLAHKMRSLLTMLGI IIGIASVSV VALNGSQKKILEDISSMGTNTISIFPGRG
orf134-1 MSVQAVLAHKMRSLLTMLGI IIGIASVSV VALNGSQKKILEDISSIGTNTISIFPGRG

orf134ng FGDRRSGKIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV
orf134-1 FGDRRSGRIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV

orf134ng RGLKLETGR LFDENDVKEDAQVVVIDQNVKDLFADSDPLGKTILFRKRPLTVIGVMKKD
orf134-1 RGLKLETGR LFDENDVKEDAQVVVIDQNVKDLFADSDPLGKTILFRKRPLTVIGVMKKD

orf134ng ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTRVAEKGLAELLKARHGTE
orf134-1 ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTQVAEKGLTDLKARHGTE

orf134ng DFFMNNSDSIRQMVESTTGTMKLLISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMA
orf134-1 DFFMNNSDSIRQIVESTTGTMKLLISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMA

orf134ng IGARRGNILQQFLIEAVLIC IIGGLVGVGLSAAVSLVFNHFVTDFFPMDIS AASVIGAVAC
orf134-1 IGARRGNILQQFLIEAVLIC IIGGLVGVGLSAAVSLVFNHFVTDFFPMDIS AMSVIGAVAC

orf134ng STGIGIAFGFMPANKAAKLNPIDALAQDX
orf134-1 STGIGIAFGFMPANKAAKLNPIDALAQDX

```

50 ORF134ng also shows homology to an *E.coli* ABC transporter:

55  
60

```

sp|P75831|YBJZ_ECOLI HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN YBJZ >gi5
(AE000189) o648; similar to YBBA_HAEIN SW: P45247 [Escherichia coli] Length =
648
Score = 297 bits (753), Expect = 6e-80
Identities = 162/389 (41%), Positives = 230/389 (58%), Gaps = 1/389 (0%)

Query: 1 MSVQAVLAHKMRSLLTMLXXXXXXXXXXXXLNGSQKKILEDISSMGTNTISIFPGRG 60
M+ +A+ A+KMR+LLTML +G+ +++ +L DI S+GTNTI ++PG+
Sbjct: 260 MAWRALAANKMRTLTLTMLGI IIGIASVSVIVVVGDAAKQMVLA DIRSIGTNTIDVYPGKD 319

Query: 61 FGDRRSGKIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV 120
FGD + L DD I KQ +VASATP S L Y N D+ AS GV YF+V
Sbjct: 320 FGDDDPQYQQALKYDDLIAIQKQPWVASATPAVSQNLRRLRYNNVDVAASANGVSGDYFNV 379

```

Query: 121 RGLKLETGRLEFDENDVKEDAQVVVIDQNVKDKLFAD-SDPLGKTLFRKRPLTVIGVMKK 179  
 G+ G F++ + AQVVV+D N + +LF +D +G+ IL P VIGV ++  
 Sbjct: 380 YGMTFSEGNTFNQEQLNGRAQVVVLDSNTRRQLFPHKADVVEVILVGNMPARVIGVAEE 439

5 Query: 180 DENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTRVAEKGLAELLKARHGT 239  
 ++ FG+S VL +W PY+T+ ++ G+S NSITV++K+ ++ AE+ L LL RHG  
 Sbjct: 440 KQSMFGSSKVLRVWLPYSTMSGRVMGQSWLNSITVRVKEGFDSAEAEQQLTRLLSLRHGK 499

10 Query: 240 EDFFMNSDSIRQMVESTTGTMKXXXXXXXXXXXXXVVGGIGVMNIMLVSVTERTKEIGIRM 299  
 +DFF N D + + VE TT T++ VVGGIGVMNIMLVSVTERT+EIGIRM  
 Sbjct: 500 KDFFTWNMDGVLKTEKTRTLQLFLTLVAVISLVVGGIGVMNIMLVSVTERTREIGIRM 559

15 Query: 300 AIGARRGNILQQFLIEXXXXXXXXXXXXXXXXXXXXXFNHFVTDPFMDISAASVIGAVA 359  
 A+GAR ++LQQFLIE F+ + + S +++ A  
 Sbjct: 560 AVGARASDVLQQFLIEAVLVCLVGGALGITLSLLIAFTLQLFLPGWEIGFSPLALLLAFL 619

20 Query: 360 CSTGIGIAFGFMPANKAAKLNPIDALAQD 388  
 CST GI FG++PA AA+L+P+DALA++  
 Sbjct: 620 CSTVTGILFGWLPARNAARLDPVDALARE 648

Based on this analysis, including the presence of the leader peptide and transmembrane regions in the gonococcal protein, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 65

25 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 539>:

1 ..GGGACGGGAG CGATGCTGCT GCTGTTTTAC GCGGTAACGA T.CTGCCTTT  
 51 GGCCACTGGC GTTACCCTGA GTTACACCTC GTCGATTTT TTGGCGGTAT  
 101 TTTCCTTCCT GATTTTGAAA GAACGGATTT CCGTTTACAC GCAGGCGGTG  
 151 CTGCTCCTTG GTTTTGCCGG CGTGGTATTG CTGCTTAATC CCTCGTTCGG  
 201 CAGCGGTCAG GAAACGGCGG CACTCGCCGG GCTGGCGGGC GGC CGCATGT  
 251 CCGGCTGGGC GTATTTGAAA GTGCGCGAAC TGTCTTTGGC GGGCGAACCC  
 301 GGCTGGCGCG TCGTGTTTTA CCTTTCGGTG ACAGGTGTGG CGATGTCGTC  
 351 GGTTTGGGCG ACGCTGACCG GCTGGCACAC CCTGTCCTTT CCATCGGCAG  
 401 TTTATCTGTC GTGCATCGGC GTGTCCGCGC TGATTGCCCA ACTGTCGATG  
 35 451 ACGCGCGCCT ACAAAGTCGG CGACAAATTC ACGGTTGCCT CGCTTTCCTA  
 501 TATGACCGTC GTTTTTTCCG CTCTGTCTGC CGCATTTTTT CTGGGCGAAG  
 551 AGCTTTTCTG GCAGGAAATA CTCGGTATGT GCATCATCAT CCTCAGCGGT  
 601 ATTTTGA

This corresponds to the amino acid sequence <SEQ ID 540; ORF135>:

40 1 ..GTGAMLLLFY AVTILPLATG VTLSTSSIF LAVFSFLILK ERISVYTQAV  
 51 LLLGFAGVVL LLNPSFRSGQ ETAALAGLAG GAMSGWAYLK VRELSLAGEP  
 101 GWRVVFYLSV TGVAMSSVWA TLTGWHTLSF PSAVYLSGIG VSALIAQLSM  
 151 TRAYKVGDKF TVASLSYMTV VFSALSAFF LGEELFWQEI LGMCIISAV  
 201 F\*

45 Further work revealed the complete nucleotide sequence <SEQ ID 541>:

1 ATGGATACCG CAAAAAAGA CATTTTAGGA TCGGGCTGGA TGCTGGTGGC  
 51 GGCGGCTGCG TTTACCATA TGAACGTATT GATTAAAGAG GCATCGGCAA  
 101 AATTGTCCTT CGGCAGCGGC GAATTGGTCT TTTGGCGCAT GCTGTTTTCA  
 50 151 ACCGTTGCGC TCGGGGCTGC CGCCGTATTG CGTCGGGACA mCTTCCGCAC  
 201 GCCCCATTGG AAAAACCCT TAAACCGCAG TATGGTCGGG ACGGGGGCGA  
 251 TGCTGTGCT GTTTTACGCG GTAACGCATC TGCTTTTGGC CACTGGCGTT  
 301 ACCCTGAGTT ACACCTCGTC GATTTTTTTG GCGGTATTTT CCTTCTGAT  
 351 TTTGAAAGAA CGGATTTCCT TTTACACGCA GCGGTGCTG CTCCTGGTT  
 401 TTGCCGCGCT GGTATTGCTG CTTAATCCCT CGTCCGCGAG CGGTCAGGAA  
 55 451 ACGGCGGCAC TCGCCGGGCT GCGGGGCGGC GCGATGTCCG GCTGGGCGTA  
 501 TTTGAAAGTG CGCGAAGTGT CTTTGGCGGG CGAACCCGGC TGGCGCGTCG  
 551 TGTTTTACCT TTCCGTGACA GGTGTGGCGA TGTCGTGCGT TTGGGCGACG

5  
601 CTGACCGGCT GGCACACCCT GTCCTTTCCA TCGGCAGTTT ATCTGTCGTG  
651 CATCGGCGTG TCCGCGCTGA TTGCCCAACT GTCGATGACG CGCGCCTACA  
701 AAGTCGGCGA CAAATTCACG GTTGCTCGC TTTCTATAT GACCGTCGTT  
751 TTTTCGCTC TGTCTGCCG ATTTTCTCTG GCGAAGAGC TTTTCTGGCA  
801 GGAAATACTC GGTATGTGCA TCATCATCCT CAGCGGTATT TTGAGCAGCA  
851 TCCGCCCCAC TGCCTTCAA CAGCGGCTGC AATCCCTGTT CCGCCAAAGA  
901 TAA

This corresponds to the amino acid sequence <SEQ ID 542; ORF135-1>:

10  
1 MDTAKKDILG SGWMLVAAAC FTIMNVLIKE ASAKFALGSG ELVFWRLFS  
51 TVALGAAAVL RRDxFRTPHW KNHLNRSVMG TGAMLLLFYA VTHLPLATGV  
101 TLSYTSSIFL AVFSFLILKE RISVYTQAVL LLGFAGVLL LNPFSRSGQE  
151 TAALAGLAGG AMSGWAYLKV RELSLAGEPG WRVVFYLSVT GVAMSSVWAT  
201 LTGWHTLSEF SAVYLSGIGV SALIAQLSMT RAYKVGDKFT VASLSYMTVV  
251 FSALSAAFFL GEELFWQEIL GMCIIILSGI LSSIRPTAFK QRLQSLFRQR  
301 \*

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF135 shows 99.0% identity over a 197aa overlap with an ORF (ORF135a) from strain A of *N.meningitidis*:

20  
orff135.pap GTGAMLLLFYAVTILPLATGVTLSTSSIF  
orff135a STVALGAAAVLRRDTRTPHWKNHLNRSVMGTGAMLLLFYAVTHLPLATGVTLSTSSIF  
25  
50 60 70 80 90 100  
orff135.pap 40 50 60 70 80 90  
LAVFSFLILKERISVYTQAVLLLGFGVLLLNPSFRSGQETAALAGLAGGAMSGWAYLK  
orff135a 110 120 130 140 150 160  
LAVFSFLILKERISVYTQAVLLLGFGVLLLNPSFRSGQETAALAGLAGGAMSGWAYLK  
30  
100 110 120 130 140 150  
orff135.pap VRELSTAGEPGWRVVFYLSVTGVAMSSVWATLTGWHTLSEFSAVYLSGIGVSALIAQLSM  
orff135a 170 180 190 200 210 220  
VRELSTAGEPGWRVVFYLSVTGVAMSSVWATLTGWHTLSEFSAVYLSGIGVSALIAQLSM  
35  
160 170 180 190 200  
orff135.pap TRAYKVGDKFTVASLSYMTVVSALSAAFFLGEELFWQEILGMCIIISAVFX  
orff135a 230 240 250 260 270 280  
TRAYKVGDKFTVASLSYMTVVSALSAAFFLAEELFWQEILGMCIIILSGILSSIRPTAF  
40  
orff135a KQRLQSLFRQRX  
290 300  
45

The complete length ORF135a nucleotide sequence <SEQ ID 543> is:

50  
1 ATGGATACCG CAAAAAAGA CATTTTAGGA TCGGCTGGA TGCTGGTGGC  
51 GGCGGCTGCT TTTACCATTA TGAACGTATT GATTAAAGAG GCATCGGCAA  
101 AATTTGCCCT CGGCAGCGGC GAATTGGTCT TTTGGCGCAT GCTGTTTCA  
151 ACCGTGCGC TCGGGGCTGC CGCCGTATTG CGTCGGGACA CCTTCCGCAC  
201 GCCCCATTGG AAAAACCCT TAAACCGCAG TATGGTCGGG ACGGGGCGA  
251 TGCTGCTGCT GTTTTACGCG GTAACGCATC TGCTTTGGC CACCGGCGTT  
301 ACCCTGAGTT ACACCTCGTC GATTTTTTGG GCGGTATTT CCTTCTGAT  
351 TTTGAAAGAA CGGATTTCCG TTTACACGCA GGCGGTGCTG CTCCTTGGTT  
55  
401 TTGCCGCGCT GGTATTGCTG CTTAATCCCT CGTCCGCGAG CGGTACAGAA  
451 ACGGCGGCAC TCGCCGGGCT GCGGGGCGGC GCGATGTCCG GCTGGGCGTA  
501 TTTGAAAGTG CGGCAACTGT CTTGGCGGG CGAACCCGCG TGCGCGCTCG  
551 TGTTTTACCT TTCCGTGACA GGTGTGGCGA TGTCATCGGT TTGGGCGACG  
601 CTGACCGGCT GGCACACCCT GTCCTTTCCA TCGGCAGTTT ATCTGTCGTG

5  
 651 CATCGGCGTG TCCGCGCTGA TTGCCCAACT GTCGATGACG CGCGCCTACA  
 701 AAGTCGGCGA CAAATTCACG GTTGCCCTCGC TTTCCTATAT GACCGTCGTT  
 751 TTTTCCGCTC TGTCTGCCGC ATTTTCTCTG GCCGAAGAGC TTTTCTGGCA  
 801 GGAAATACTC GGTATGTGCA TCATCATCCT CAGCGGTATT TTGAGCAGCA  
 851 TCCGCCCCAC TGCCTTCAAA CAGCGGCTGC AATCCCTGTT CCGCCAAAGA  
 901 TAA

This encodes a protein having amino acid sequence <SEQ ID 544>:

10  
 1 MDTAKKDILG SGWMLVAAAC FTIMNVLIKE ASAKFALGSG ELVFWRLFS  
 51 TVALGAAAVL RRDTRTPHW KNHLNRSMSG TGAMLLLFYA VTHLPLATGV  
 101 TLSYTSSIFL AVFSFLILKE RISVYTQAVL LLGFAGVLL LNPSFRSGQE  
 151 TAALAGLAGG AMSGWAYLKV RELSLAGEPG WRVVFYLSVT GVMSSVWAT  
 201 LTGWHTLSFP SAVYLSGIGV SALIAQLSMT RAYKVGDKFT VASLSYMTVV  
 251 FSALSAAFFL AEELFWQEIL GMCIIILSGI LSSIRPTAFK QRLQSLFRQR  
 301 \*

15 ORF135a and ORF135-1 show 99.3% identity in 300 aa overlap:

20  
 orf135a.pep MDTAKKDILGSGWMLVAAACFTIMNVLIKEASAKFALGSGELVFWRLFSSTVALGAAAVL  
 orf135-1 MDTAKKDILGSGWMLVAAACFTIMNVLIKEASAKFALGSGELVFWRLFSSTVALGAAAVL  
 25  
 orf135a.pep RRDTRTPHWKNHLNRSMSGTGAMLLLFYAVTHLPLATGVTLSTSSIFLAVFSFLILKE  
 orf135-1 RRDTRTPHWKNHLNRSMSGTGAMLLLFYAVTHLPLATGVTLSTSSIFLAVFSFLILKE  
 30  
 orf135a.pep RISVYTQAVLLLGAGVLLLNPSFRSGQETAALAGLAGGAMSGWAYLKVRELSLAGEPG  
 orf135-1 RISVYTQAVLLLGAGVLLLNPSFRSGQETAALAGLAGGAMSGWAYLKVRELSLAGEPG  
 35  
 orf135a.pep WRVVFYLSVTGVMSSVWATLTGWHTLSFSAVYLSGIGVSALIAQLSMT RAYKVGDKFT  
 orf135-1 WRVVFYLSVTGVMSSVWATLTGWHTLSFSAVYLSGIGVSALIAQLSMT RAYKVGDKFT  
 40  
 orf135a.pep VASLSYMTVVFSALSAAFFLAEEELFWQEILGMCIIILSGILSSIRPTAFKQRLQSLFRQR  
 orf135-1 VASLSYMTVVFSALSAAFFLGEELFWQEILGMCIIILSGILSSIRPTAFKQRLQSLFRQR

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF135 shows 97% identity over a 201aa overlap with a predicted ORF (ORF135ng) from *N.gonorrhoeae*:

40  
 orf135.pep GTGAMLLLFYAVTXLPLATGVTLSTSSIF 30  
 orf135ng STVTLGAAAVLRRDTRTPHWKNHLNRSMSGTGAMLLLFYAVTHLPLTTGVTLSTSSIF 335  
 45  
 orf135.pep LAVFSFLILKERISVYTQAVLLLGAGVLLLNPSFRSGQETAALAGLAGGAMSGWAYLK 90  
 orf135ng LAVFSFLILKERISVYTQAVLLLGAGVLLLNPSFRSGQEPALAGLAGGAMSGWAYLK 395  
 50  
 orf135.pep VRELSLAGEPGWRVVFYLSVTGVMSSVWATLTGWHTLSFSAVYLSGIGVSALIAQLSM 150  
 orf135ng VRELSLAGEPGWRVVFYLSATGVMSSVWATLTGWHTLSFSAVYLSGIGVSALIAQLSM 455  
 55  
 orf135.pep TRAYKVGDKFTVASLSYMTVVFSALSAAFFLGEELFWQEILGMCIIISAVF 201  
 orf135ng TRAYKVGDKFTVASLSYMTVVFSALSAAFFLGEELFWQEILGMCIIISAAF 506

55 An ORF135ng nucleotide sequence <SEQ ID 545> was predicted to encode a protein having amino acid sequence <SEQ ID 546>:

1 MPSEKAFRRH LRTASFQGLH LHHFHQKVCK CGIIGFGIHI FPTLLPAAQG  
 51 ILDIQLGLFR IDFAALAVYR RTQVDFIHTV IDGIASDQAF SEVVQILRRL  
 101 NLGHFTDTHL IAQARRFIAD FGNIRPMRRG EAKTFRCRCR FDGIDGIHGD

151 FRQCGHINRL APGKDCRNGK RDKVFFHTRH YNQVCLEKTN CSARKIKFRH  
 201 QKQAKTHSTS LAARFTIRPS LSQRPFMDTA KKDILGSGWM LVAAACFTVM  
 251 NVLIKEASAK FALGSGELVF WRMLFSTVTL GAAAVLRRDT FRTPHWKNHL  
 301 NRSVMVGTGAM LLLFYAVTHL PLTTGVTLSY TSSIFLAVFS FLILKERISV  
 5 YTAQVLLLG F AGVVLLLNPS FRSGQEPAL AGLAGGAMSG WAYLKVRELS  
 351 LAGEPGWRVV FYLSATGVAM SSVWATLTGW HTLSFPSAVY LSGIGVSALI  
 401 AQLSMTRAYK VGDKFTVASL SYMTVVFSAL SAAFFLGEEL FWQEILGMCI  
 451 IISAAF\*

Further work revealed the following gonococcal sequence <SEQ ID 547>:

10 1 ATGGATACCG CAAAAAAGA CATTTTAGGA TCGGGCTGGA TGCTGGTGGC  
 51 GCGCGCCTGC TTCACCGTTA TGAACGTATT GATTAAAGAG GCATCGGCAC  
 101 AATTTGCCCT CGGCAGCGGC GAATTGGTCT TTTGGCGCAT GCTGTTTTCA  
 151 ACCGTTACGC TCGGTGCTGC CGCCGTATTG CGCGCGACA CCTTCCGCAC  
 201 GCGCCATTGG AAAAACCCT TAAACCGCAG TATGGTCGGG ACGGGGGCGA  
 15 TGCTGCTGCT GTTTTACGCG GTAACGCATC TGCCTTTGAC AACCGCGCTT  
 301 ACCCTGAGTT ACACCTCGTC GATTTTTttg GCGGTATTTT CCTTCCTGAT  
 351 TTTGAAAGAA CGGATTTCOG TTTACACGCA GCGGTGCTG CTCCTTGGTT  
 401 TTGCCGGCGT GGTATTGCTG CTTAATCCCT CGTTCGCAG CGGTCAGGAA  
 451 CCGCGCGCAC TCGCCGGGCT GCGGGGCGGC GCATGTCCG GCTGGGCGTA  
 20 TTTGAAAGTG CGCGAAGTGT CTTGGCGGG CGAACCGGC TGGCGCGTCG  
 551 TGTTCCTTACCT TTCCGCAACC GCGGTGCGCA TGTCGTGgt ttggcgacg  
 601 Ctgaccggct ggCACaccT GTCCTTTcca tcggcagttt ATCgtCGGG  
 651 CATCGGCGTG tccgcgCtG TGGCCCAaCT GtcgatgAcg cGCGcctaca  
 701 aaGTCGCGCA CAAATTACG GTTGCCTCGC tttcctaTat gaccgtcGTC  
 25 TTTTCCGCCC TGTCTGCCGC ATTTTTCTg ggcgaagagc tttTctggCA  
 801 GGAAATACTC GGTATGTGCA TCATTatcT CAGCGGCATT TTGAGCAGCA  
 851 TCCGCCCAT TGCCTTCAA CAGCGGCTGC AAGCCCTCTT CCGCAAAGA  
 901 TAA

This corresponds to the amino acid sequence <SEQ ID 548; ORF135ng-1>:

30 1 MDTAKKDILG SGWMLVAAAC FTMNVLIKE ASAKFALGSG ELVFWRLFS  
 51 TVTLGAAAVL RRDTFRTPHW KNHLNRSVMG TGAMLLLFYA VTHLPLTTGV  
 101 TLSYTSIFL AVFSFLILKE RISVYTQAVL LLGFAGVLL LNPFSRSGQE  
 151 PAALAGLAGG AMSGWAYLKV RELSLAGEPG WRVVFYLSAT GVAMSSVWAT  
 201 LTGWHTLSFP SAVYLSGIGV SALIAQLSMT RAYKVGDKFT VASLSYMTV  
 35 FSALSAFFL GEELFWQEIL GMCIIILSGI LSSIRPIAFK QRLQALFRQR  
 301 \*

ORF135ng-1 and ORF135-1 show 97.0% identity in 300 aa overlap:

orf135ng-1.pep MDTAKKDILGSGWMLVAAACFTMVNVLKEASAKFALGSGELVFWRLFSTVTLGAAAVL  
 40 orf135-1 MDTAKKDILGSGWMLVAAACFTIMNVLIKEASAKFALGSGELVFWRLFSTVALGAAAVL  
 orf135ng-1.pep RRDTFRTPHWKNHLNRSVMGTGAMLLLFYAVTHLPLTTGVTLSYTSSIFLAVFSFLILKE  
 45 orf135-1 RDXFRTPHWKNHLNRSVMGTGAMLLLFYAVTHLPLATGVTLSYTSSIFLAVFSFLILKE  
 orf135ng-1.pep RISVYTQAVLLLG FAGVLLLNPSFRSGQEPALAGLAGGAMSGWAYLKVRELSLAGEPG  
 orf135-1 RISVYTQAVLLLG FAGVLLLNPSFRSGQETAALAGLAGGAMSGWAYLKVRELSLAGEPG  
 50 orf135ng-1.pep WRVVFYLSATGVAMSSVWATLTGWHTLSFPSAVYLSGIGVSALIAQLSMTRAYKVGDKFT  
 orf135-1 WRVVFYLSVTGVAMSSVWATLTGWHTLSFPSAVYLSGIGVSALIAQLSMTRAYKVGDKFT  
 orf135ng-1.pep VASLSYMTVVFSALSAFFLGEELFWQEILGMCIILSGILSSIRPIAFKQRLQALFRQR  
 55 orf135-1 VASLSYMTVVFSALSAFFLGEELFWQEILGMCIILSGILSSIRPTAFKQRLQSLFRQR

Based on this analysis, including the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

**Example 66**

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 549>:

```

      1 ATGAAGCGGC GTATAGCCGT CTTCGTCTCG TTCCCGCAGA TAATCCGAGT
      51 TTTGGGACAA CTGTTGCCGA AAATCGTCAA TACAGTTCCG GCACATCGGA
1001 TGCTCTTCCA GATTTTCGGG ATGTTCTTTT TCTTCATACA CCAGCAATAT
151 CTGCCCGGGA TCGCCGAAAT CGATTCCCCA TCGGGCATCG TGTTCCGTGC
201 GCTCCTCTTC CGTCATCTGC CCGCGCATTG CCTGTATGGT AAAGCCGCCG
251 TAGGGGATGC CgTTGCACAC GAACATCCAG TCGCTGATGT CGTCAACCGG
301 AACGCAAACG cTTTCGCCTT GTTCGACATT GGTCAGTTCG CCsGGTTCAT
10 351 TGTTCAGCAC ACCGTAAATA TAAAGACCGT CAAAATAAAT ATCGTTCGATC
401 CACATATGTT CGCAAATTC GCGTCTTCG CCGTCTTGGA AAAAAGGGAC
451 TTTGACCATG GCAAAATCCA AGCGGAAAT AATGCGGCGG CGTTCCTCAA
501 AAAGcTCGCG CCAAAAATAT TTGAATGTTT TACGGGCGCG TTGTCGCGCA
551 CGGTTTACCG GTTCGTCTGC CTGTTCTACA TAATAAATGA CGGAATCGCC
15 601 CATCATATCT GCTCCTCAAC GTGTACGTA TGTGTTTGA CCTTACTCGG
651 GCTTCTTgCc kTCGGCATCC GATTGGGATT TGAAAAGTTC mmrwyATTTCG
701 GAATAG

```

This corresponds to the amino acid sequence <SEQ ID 550; ORF136>:

```

      1 MKRRIAEVFL FPQIIRVLGQ LLPKIVNTVP AHRMLFQIFG MFFFFFIHQY
      51 LPGIAEIDSP CGIVFGALLF RHLPAHCLYG KAAVGDAVAH EHPVADVNR
20 101 NANAFALFDI GGFAXFIVQH TVNIKTVKIN IVDPHMFANF AVFAVLEKRD
151 FDHGKIQGGN NAAAFPKKLA PKIFECFTGA FVGTVYRFVC LFYIINDGIA
201 HHSAPQVRVY LFPAPYCGFL PSASDSDLKSS XXSE*

```

Further work revealed the complete nucleotide sequence <SEQ ID 551>:

```

      1 ATGATGAAGC GGCGTATAGC CGTCTTCGTC CTGTTCCCGC AGATAATCCG
      51 AGTTTTGGGA CAACTGTTGC CGAAAATCGT CAATACAGTT CCGGCACATC
25 101 GGATGCTCTT CCAGATTTTC GGGATGTTCT TTTTCTTCAT ACACCAGCAA
151 TATCTGCCCG GGATCGCCGA AATCGATTCC CCATGCGGCA TCGTGTTCGG
201 TGCGCTCCTC TTCCGTCATC TGCCCGCGCA TTGCCTGTAT GGTAAGCCG
30 251 CCGTAGGGGA TGCCGTTGCA CACGAACATC CAGTCGCTGA TGTCGTCAAC
301 CGGAACGCAA ACGCTTTCGC CTGTTTCGAC ATTGGTCAGT TCGCCGGGTT
35 351 CATTGTTTCA CACACCGTAA ATATAAGAC CGTCAAAATA AATATCGTCG
401 ATCCACATAT GTTCGCAAAT TTCGCCGTCT TCGCCGTCTT GGAAAAAAGG
451 GACTTTTGACC ATGGCAAAAT CCAAGGCGGA AATAATGCGG CGGCGTTCCC
15 501 AAAAAAGCTC GCGCAAAAAA TATTGAATG TTTTACGGGC GCGTTCGTCG
551 GCACGGTTTA CCGGTTTCGTC TGCTGTCTT ACATAATAAA TGACGGAATC
601 GCCCATCATT CTGCTCCTCA ACGGTACGG TATCTGTTTG CACCTTACTG
651 CGGCTTTCTG CCTTCGGCAT CCGATTCCGA TTTGAAAAGT TCCAAATATT
701 CGGAATAG

```

This corresponds to the amino acid sequence <SEQ ID 552; ORF136-1>:

```

      1 MMKRIAEVFL FPQIIRVLG QLLPKIVNTV PAHRMLFQIF GMFFFFIHQQ
      51 YLPGIAEIDS PCGIVFGALL FRHLPAHCLY GKAAGDAVA HEHPVADVNR
45 101 RNANAFALFD IGQFAGFIVQ HTVNIKTVKI NIVDPHMFAN FAVFAVLEKR
151 DFDHGKIQGG NNAAFPKKLA APKIFECFTG AFVGTVYRFV CLFYIINDGI
201 AHHSAPQVRV YLFPYCGFL PSASDSDLKS SKYSE*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF136 shows 71.7% identity over a 237aa overlap with an ORF (ORF136a) from strain A of *N. meningitidis*:

```

50      10      20      30      40      50      59
orf136.pep MKRRIAEVFLFPQIIRVLGQLLPKIVNTVPAHRMLFQIFGMFFFFFIHQYLPGLAEIDS
          |||||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf136a    MMKRIAEVLLMQKIRILGQLLPKIVNTVPAHRMLFQXFGMFFFFFIHQYLPGLAEIDS
          10      20      30      40      50      60

```





190 200 210 220 230

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF136 shows 92.3% identity over a 234aa overlap with a predicted ORF (ORF136ng) from

#### 5 *N.gonorrhoeae*:

orf136.pep	MKRRIAVFVLFPPQIIIRVLGQLLPKIVNTVPAHRMLFQIFGMFFFFFIHQYLPGLAEIDS	59
orf136ng	MMKRRIAVFVLLMQKIRILGQLLPKIVNTVPAHRMLFQIFGMFFFFFIHQYLPGLAEIDS	60
orf136.pep	PCGIVFGALLFRHLPAHCLYGKAAVGDVAHEHPVADVNNANAFALFDIGQFAXFIVQ	119
orf136ng	PGGIVFGTLLFRHLSAHCLYGKAAVGDVAHEHPVADVNNANAFALFDIGQSAGFIVQ	120
orf136.pep	HTVNIKTVKINIVDPHMFANFAVFAVLEKRDFDHGKIQGGNNAAFPKKLAPKIFECFTG	179
orf136ng	HTVNIKTVKINIVDPHMFANFAVFAVLEKRDFDHGKIQGGNNAAFPKKLAPKVFECFTG	180
orf136.pep	AFVGTVYRFVCLFYIINDGIAHHSAPQVRVYLFAPYCGFLPSASDSLKSSXXSE	234
orf136ng	AFAGTVYRFVCLFYIINDGIAHHTAPQVRVYLFAPYRGFLPPASDSLKSSKYSE	235

The complete length ORF136ng nucleotide sequence <SEQ ID 555> is:

1	ATGATGAAGC	GGCGTATAGC	CGTCTTCGTC	CTGCTCATGC	AGAAAATCCG
51	GATTTTGGGA	CAACTGTTGC	CGAAAATCGT	CAATACAGTT	CCGGCACATC
101	GGATGCTCTT	CCAAATTTTC	GGGATGTTCT	TTTCTTCAT	ACACCGGCAA
151	TACCTGCCCG	GGATCGCCGA	AATCGATTCC	CCAGGCGGTA	TCGTGTTCGG
201	TACGCTCCTC	TTCCGTCATC	TGTCCGCGCA	TTGCCTGTAC	GGTAAAGCCG
251	CCGTAGGGGA	TGCCGTTGCA	CACGAACATC	CAGTCGCTGA	TGTCGCCAAC
301	CGGAACGCAA	ACGCTTTTCG	CTTGTTTCGAC	ATTGGTCAGT	CCGCCGGGTT
351	CATTGTTCAG	CACACCGTAA	ATATAAGAC	CGTCAAAATA	AATATCGTCG
401	ATCCACATAT	GTTCGCAAAAT	TTCCGCGTCT	TCGCCGTCTT	GGAAAAAAGG
451	GACTTTGACC	ATGGCAAAAT	CCAAGGCGGA	AATAATGCGG	CGCGTTTCCC
501	AAAAAAGCTC	GCGCCAAAAG	TATTTGAATG	TTTACGGGC	GCGTTCGCCG
551	GCACGGTTTA	CCGGTTCGTC	TGCCTGTTCT	ACATAATAAA	TGACGGAATC
601	GCCCATCATA	CTGCTCCTCA	ACGTGTACGG	TATCTGTTTG	CACCTTACCG
651	CGGTTTTCTA	CCTCCGGCAT	CCGATTCGGA	TTTGAAAAGT	TCCAAATATT
701	CGGAATAG				

This encodes a protein having amino acid sequence <SEQ ID 556>:

1	MMKRRIAVFV	LLMQKIRILG	QLLPKIVNTV	PAHRMLFQIF	GMFFFFIHRQ
51	YLPGLAEIDS	PGGIVFGTLL	FRHLSAHCLY	GKAAVGDVA	HEHPVADVAN
101	RNANAFALFD	IGQSAGFIVQ	HTVNIKTVKI	NIVDPHMFAN	FAVFAVLEKR
151	DFDHGKIQGG	NNAAFPKKL	APKVFECFTG	AFAGTVYRFV	CLFYIINDGI
201	AHHTAPQVRV	YLFAPYRGFL	PPASDSLKSS	SKYSE*	

ORF136ng and ORF136-1 show 93.6% identity in 235 aa overlap:

orf136ng	MMKRRIAVFVLLMQKIRILGQLLPKIVNTVPAHRMLFQIFGMFFFFFIHQYLPGLAEIDS	
orf136-1	MMKRRIAVFVLFPPQIIIRVLGQLLPKIVNTVPAHRMLFQIFGMFFFFFIHQYLPGLAEIDS	
orf136ng	PGGIVFGTLLFRHLSAHCLYGKAAVGDVAHEHPVADVNNANAFALFDIGQSAGFIVQ	
orf136-1	PCGIVFGALLFRHLPAHCLYGKAAVGDVAHEHPVADVNNANAFALFDIGQFAGFIVQ	
orf136ng	HTVNIKTVKINIVDPHMFANFAVFAVLEKRDFDHGKIQGGNNAAFPKKLAPKVFECFTG	
orf136-1	HTVNIKTVKINIVDPHMFANFAVFAVLEKRDFDHGKIQGGNNAAFPKKLAPKIFECFTG	
orf136ng	AFAGTVYRFVCLFYIINDGIAHHTAPQVRVYLFAPYRGFLPPASDSLKSSKYSEX	
orf136-1	AFVGTVYRFVCLFYIINDGIAHHSAPQVRVYLFAPYCGFLPSASDSLKSSKYSEX	

Based on the presence of the putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 67

5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 557>:

```

1  ATGGAAAATA TGGTAACGTT TTCAAAAATC AGACCGCTTT TGGCAATCGC
51  CGCCGCCGCG TTGCTTGCCG CC.TGCGGAC GGCGGGAAAT AATGCTGTCC
101 GCAAGCCGGT GCAAACCGCC AAACCCGCCG CAGTGGTCGG TTTGGCACTC
151 GGTGGCGGCG CATCTAAAGG ATTTGCCCAT GTAGGTATTA TTAAGGTTTT
201 GAAAGAAAAC GGTATTCCTG TGAAGGTGGT TACCGGCACC TCCGCAGGTT
251 CGATTGTCGG CAACCTTTT GCATCGGGTA TGTCGCCCGA CCGCCTCGAA
301 TTGGAAGCCG AAATTTTAGG CAAAACCGAT TTGGTCGATT TAACCTTGTC
351 CACCAATGGG TTATCAAAG GCGCAAAGCT GCAAAATAC ATCAACCGAA
401 AACTCCGCGG CATGCAGATT CAGCAGTTTC CCATCAAATT TGCCGCC..

```

15 This corresponds to the amino acid sequence <SEQ ID 558; ORF137>:

```

1  MENMVTFSKI RPLLAIAAAA LLAAXRTAGN NAVRKPVQTA KPAAVVGGLAL
51  GGGASKGFAH VGIIKVLKEN GIPVKVVTGT SAGSIVGNLF ASGMSPDRLE
101 LEAEILGKTD LVDLTLSTNG FIKGAKLQNY INRKLGRMQI QQFPIKFAA..

```

Further work revealed the complete nucleotide sequence <SEQ ID 559>:

```

1  ATGGAAAATA TGGTAACGTT TTCAAAAATC AGACCGCTTT TGGCAATCGC
51  CGCCGCCGCG TTGCTTGCCG CCTGCGGCAC GGCGGGAAAT AATGCTGTCC
101 GCAAGCCGGT GCAAACCGCC AAACCCGCCG CAGTGGTCGG TTTGGCACTC
151 GGTGGCGGCG CATCTAAAGG ATTTGCCCAT GTAGGTATTA TTAAGGTTTT
201 GAAAGAAAAC GGTATTCCTG TGAAGGTGGT TACCGGCACA TCGGCAGGTT
251 CGATTGTCGG CAGCCTTTT GCATCGGGTA TGTCGCCCGA CCGCCTCGAA
301 TTGGAAGCCG AAATTTTAGG CAAAACCGAT TTGGTCGATT TAACCTTGTC
351 CACCAAGTGGT TTTATCAAAG GCGAAAAGCT GCAAAATTAC ATCAACCGAA
401 AAGTCGGCGG CAGGCAGATT CAGCAGTTTC CCATCAAATT TGCCGCCGTT
451 GCTACTGATT TTGAAACCGG CAAGGCCGTC GCTTCAATC AGGGGAATGC
501 CGGGCAGGCT GTGCGCGCTT CCGCCGCCAT TCCCAATGTG TTCCAACCGG
551 TTATCATCGG CAGGCATACA TATGTTGACG GCGGTCTGTC GCAGCCCGTG
601 CCCGTCAAGT CCGCCCGGCG GCAGGGGGCG AATTTCGTGA TTGCCGTCGA
651 TATTTCCGCC CGTCCGGGCA AAAACATCAG CCAAGGTTTC TTCTCTTATC
701 TCGATCAGAC GCTGAACGTA ATGAGCGTTT CTGCGTTGCA AAATGAGTTG
751 GGGCAGGCGG ATGTGGTTAT CAAACCGCAG GTTTTGGATT TGGGTGCAGT
801 CGGCGGATTC GATCAGAAAA AACGCGCCAT CCGGTGGGGT GAGGAGGCAG
851 CACGTGCCGC ATTGCTGAA ATCAAACGCA AACTGGCGGC ATACCGTTAT
901 TGA

```

This corresponds to the amino acid sequence <SEQ ID 560; ORF137-1>:

```

1  MENMVTFSKI RPLLAIAAAA LLAACGTAGN NAVRKPVQTA KPAAVVGGLAL
51  GGGASKGFAH VGIIKVLKEN GIPVKVVTGT SAGSIVGSLF ASGMSPDRLE
101 LEAEILGKTD LVDLTLSTSG FIKGEKLQNY INRKVGGRQI QQFPIKFAAV
151 ATDFETGKAV AFNQGNAQQA VRASAAIPNV FQPVIIGRHT YVDGGLSQPV
201 PVSAARRQGA NFVIAVDISA RPKNISQGF FSYLDQTLNV MSVSALQNEL
251 GQADVVIKQP VLDLGAVGGF DQKKRAIRLG EEAARAALPE IKRKLAAARY
301 *

```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF137 shows 93.3% identity over a 149aa overlap with an ORF (ORF137a) from strain A of *N.*

50 *meningitidis*:

```

      10      20      30      40      50      60
orf137.pep  MENMVTFSKIRPLLAIAAAALLAAXRTAGNNAVRKPVQTAKPAAVVGLALGGGASKGFAH
      |||||:|||||
orf137a     MENMVTFSKIRPLLAIAAAALLAACGTAGNNAARKPVQTAKPAAVVGLALGGGASKGFAH
      10      20      30      40      50      60

      70      80      90     100     110     120
orf137.pep  VGIKVLKENGIPVKVVTGTSAGSIVGNLFASGMSPDRLELEAEILGKTDLVDLTLSTNG
      |||||:|||||
orf137a     VGIKVLKENGIPVKVVTGTSAGSIVGSLFASGMSPDRLELEAEILGKTDLVDLTLSTSG
      70      80      90     100     110     120

      130     140     149
orf137.pep  FIKGAKLQNYINRKLGRMQIQFPIKFAA
      ||| |||||:|
orf137a     FIKGEKLQNYINRKVGGRRIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV
      130     140     150     160     170     180

```

The complete length ORF137a nucleotide sequence <SEQ ID 561> is:

```

      1  ATGGAAAATA  TGGTAACGTT  TTCAAAAATC  AGACCGCTTT  TGGCAATCGC
20  51  CGCGCCGCGC  TTGCTTGCCG  CCTGCGGCAC  GGCGGGAAT  AATGCTGCCC
101  GCAAGCCGGT  GCAAACCGCC  AAACCCGCGC  CAGTGGTCGG  TTGGCACTC
151  GGTGGCGGCG  CATCTAAAGG  ATTTGCCCAT  GTAGGTATTA  TTAAGGTTTT
201  GAAAGAAAAC  GGTATTCCTG  TGAAGGTGGT  TACCGGCACA  TCGGCAGGTT
25  251  CGATAGTCGG  CAGCCTTTTT  GCATCGGGTA  TGTCGCCGA  CCGCCTCGAA
301  TTGGAAGCCG  AAATTTTAGG  TAAAACCGAT  TTGGTCGATT  TAACCTTGTC
351  CACCAAGTGT  TTTATCAAAG  GCGAAAAGCT  GCAAAATTAC  ATCAACCGAA
401  AAGTCGGCGG  CAGGCGGATT  CAGCAGTTTC  CCATCAAATT  TGCCGCCGTT
451  GCTACTGATT  TTGAAACCGG  CAAGGCCGTC  GCTTCAATC  AAGGGAATGC
501  CGGGCAGGCT  GTGCGCGCTT  CCGCCGCCAT  TCCCAATGTG  TTCCAACCGC
30  551  TTATCATCGG  CAGGCATACA  TATGTTGACG  GCGGTCTGTC  GCAGCCCGTG
601  CCCGTCAGTG  CCGCCCGGCG  GCANGNNNG  NATNTCGTGA  TTGCCGTCGA
651  TATTTCCGCC  CGTCCGAGCA  AAAACATCAG  CCAAGGCTTC  TTCTCTTATC
701  TCGATCAGAC  GCTGAACGTA  ATGAGCGTTT  CCGCGTTGCA  AAATGAGTTG
751  GGCAGGCGG  ATGTGGTTAT  CAAACCGCAG  GTTTTGGATT  TGGGTGCAGT
35  801  CGGCGGATTC  GATCAGAAAA  AACGCGCCAT  CCGGTTGGGT  GAGGAGGCAG
851  CACGTGCCGC  ATTGCCTGAA  ATCAAACGCA  AACTGGCGGC  ATACCGTTAT
901  TGA

```

This encodes a protein having amino acid sequence <SEQ ID 562>:

```

      1  MENMVTFSKI  RPLLAIAAAA  LLAACGTAGN  NAARKPVQTA  KPAAVVGLAL
40  51  GGGASKGFAH  VGIKVLKEN  GIPVKVVTGT  SAGSIVGSLF  ASGMSPDRLE
101  LEAEILGKTD  LVDLTLSTG  FIKGEKLQNY  INRKVGGRRI  QQFPIKFAAV
151  ATDFETGKAV  AFNQGNAGQA  VRASAAIPNV  FQPVIIGRHT  YVDGGLSQPV
201  PVSAARRXXX  XXVIAVDISA  RPSKNISQGF  FSYLDQTLNV  MSVSALQNEL
251  QQADVVIKPP  VLDLGA VGGF  DQKKRAIRLG  EEARAALPE  IKRKLAAARY
45  301  *

```

ORF137a and ORF137-1 show 97.3% identity in 300 aa overlap:

```

orf137a.pep  MENMVTFSKIRPLLAIAAAALLAACGTAGNNAARKPVQTAKPAAVVGLALGGGASKGFAH
      |||||:|||||
orf137-1     MENMVTFSKIRPLLAIAAAALLAACGTAGNNAVRKPVQTAKPAAVVGLALGGGASKGFAH

orf137a.pep  VGIKVLKENGIPVKVVTGTSAGSIVGSLFASGMSPDRLELEAEILGKTDLVDLTLSTSG
      |||||:|||||
orf137-1     VGIKVLKENGIPVKVVTGTSAGSIVGSLFASGMSPDRLELEAEILGKTDLVDLTLSTSG

orf137a.pep  FIKGEKLQNYINRKVGGRRIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV
      |||||:|||||
orf137-1     FIKGEKLQNYINRKVGGRRIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV

orf137a.pep  FQPVIIGRHTYVDGGLSQPVVSAARRXXXXXVIAVDISARPSKNISQGFSSYLDQTLNV
      |||||:|||||
orf137-1     FQPVIIGRHTYVDGGLSQPVVSAARRQGANFVIAVDISARPGKNISQGFSSYLDQTLNV

orf137a.pep  MSVSALQNELGQADVVIKPPVLDLGA VGGFDQKKRAIRLGEEAARAALPEIKRKLAAARY

```

orfl37-1 MSVSALQNELGQADVVIKPVQLDLGAVGGFDQKKRAIRLGEEAARAALPEIKRKLAAARY

### Homology with a predicted ORF from *N.gonorrhoeae*

- 5 ORF137 shows 89.9% identity over a 149aa overlap with a predicted ORF (ORF137ng) from *N.gonorrhoeae*:

orfl37.pep	MENMVTFSKIRPLLAIAAAAALLAAXRTAGNNAVRKPVQTAKPAAVVGLALGGGASKGFAH	60
orfl37ng	MENMVTFSKIRSFLAIAAAAALLAACGTAGNNAARKPVQTAKPAAVVALALGGGASKGFAH	60
orfl37.pep	VGIIKVLKENGIPVKVVTGTSAGSIVGNLFASGMSDRLELEAEILGKTDLVDLTLSTNG	120
orfl37ng	IGIVKVLKENGIPVKVVTGTSAGSIVGSLASGMSDRLELEAEILGKTDLVDLTLSTSG	120
orfl37.pep	FIKGAQLQNYINRKLGMQIQFPIKFAA	149
orfl37ng	FIKGEKLQNYINRKVGGRQIQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV	180

The complete length ORF137ng nucleotide sequence <SEQ ID 563> is:

1	ATGGAAAATA	TGGAACGTT	TTCAAAAATC	AGATCATTTT	TGGCAATCGC
51	CGCCGCCGCG	TTGCTTGCCG	CCTGCGGTAC	GGCGGGAAC	AATGCCGCCC
101	GCAAGCCGGT	GCAAACCGCC	AAACCCGCGC	CAGTGGTCGC	TTTGGCACTC
151	GGTGGCGGCG	CATCTAAAGG	ATTGCCCCAT	ATAGGAATTG	TTAAGGTTTT
201	GAAAGAAAAC	GGTATTCCTG	TGAAGGTGGT	TACCGGCACA	TCGGCAGGTT
251	CGATAGTCGG	CAGCCTTTTG	GCATCGGGTA	TGTCGCCCGA	CCGCCTCGAA
301	TTGGAAGCCG	AGATTTTAGG	TAAAACCGAT	TTAGTCGATT	TACCTTGTC
351	CACCACTGGT	TTTATCAAAG	GCGAAAAGCT	GCAAAATTAC	ATCAACCGAA
401	AAGTCGGCGG	CAGGCAGATT	CAGCAGTTTC	CCATCAAATT	TGCCGCCGTT
451	GCCACTGATT	TTGAAACCGG	CAAGGCCGTC	GCTTCAATC	AAGGGAATGC
501	CGGGCAGGCG	GTTCGTGCTT	CCGCCGCCAT	TCCCAATGTG	TTCCAGCCAG
551	TCATCATCGG	CAGGCACAAA	TATGTTGACG	GCGGTCTGTC	GCAGCCCGTG
601	CCCGTCAGTG	CCGCTCGGCG	GCAGGGGGCG	AATTCGTGA	TTGCCGTCGA
651	TATTTCCGCA	CGTCCGAGCA	AAAATGTCGG	TCAAGGTTTC	TTCTCTTATC
701	TCGATCAGAC	GCTGAACGTG	ATGAGCGTTT	CCGTGTTGCA	AAACGAGTTG
751	gggcAGGCGG	ATGTGGTTAT	CAAACCGCag	gtTTTGGATT	TGGGTGCAGT
801	CGGCGGATTC	GATCAGAAAA	AGCGCGCCAT	CCGTTTGGGC	GAGGAGGCAG
851	CACGTGCCGC	ATTGCCTGAA	ATCAAACGCA	AACTGGCGGC	ATACCGTTAT
901	TGA				

This encodes a protein having amino acid sequence <SEQ ID 564>:

1	MENMVTFSKI	RSFLAIAAAA	LLAACGTAGN	NAARKPVQTA	KPAAVVALAL
51	GGGASKGFAH	IGIVKVLKEN	GIPVKVVTGT	SAGSIVGSL	ASGMSDRLE
101	LEAEILGKTD	LVDLTLSTSG	FIKGEKLQNY	INRKVGGRQI	QQFPIKFAAV
151	ATDFETGKAV	AFNQGNAGQA	VRASAAIPNV	FQPVIIGRHK	YVDGGLSQPV
201	PVSAARRQGA	NFVIAVDISA	RPSKNVGQGF	FSYLDQTLNV	MSVSVLQNEL
251	GQADVVIKPO	VLDLGAAGGF	DQKKRAIRLG	EEAARAALPE	IKRKLAAARY
301	*				

ORF137ng and ORF137-1 show 96.0% identity in 300 aa overlap:

orfl37ng	MENMVTFSKIRSFLAIAAAAALLAACGTAGNNAARKPVQTAKPAAVVALALGGGASKGFAH
orfl37-1	MENMVTFSKIRPLLAIAAAAALLAACGTAGNNAVRKPVQTAKPAAVVGLALGGGASKGFAH
orfl37ng	IGIVKVLKENGIPVKVVTGTSAGSIVGSLASGMSDRLELEAEILGKTDLVDLTLSTSG
orfl37-1	VGIIKVLKENGIPVKVVTGTSAGSIVGSLFASGMSDRLELEAEILGKTDLVDLTLSTSG
orfl37ng	FIKGEKLQNYINRKVGGRQIQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV
orfl37-1	FIKGEKLQNYINRKVGGRQIQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV
orfl37ng	FQPVIIGRHKYVDGGLSQPVVSAARRQGANFVIAVDISARPSKNVGQGFSSYLDQTLNV

orf137-1 FQPVIIGRHTYVDGGLSQPVPVSAARRQGANFVIAVDISARPGKNISQGFSSYLDQTLNV  
 orf137ng MSVSVLQNELGQADVVIKPOVLDLGAVGGFDQKKRAIRLGEEAARAALPEIKRKLAAYRY  
 ||||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 5 orf137 MSVSALQNELGQADVVIKPOVLDLGAVGGFDQKKRAIRLGEEAARAALPEIKRKLAAYRY

Based on the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site (underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

## 10 Example 68

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 565>:

1 ATGTTTCGTT TACAATTCAG GCTGTTTCCC CCTTTGCGAA CCGCCATGCA  
 51 CATCCTGTTG ACCGCCCTGC TCAAATGCCT CTCCTGCTG CCGCTTTCCT  
 101 GTCTGCACAC GCTGGGAAAC CGGCTCGGAC ATCTGGCGTT TTACCTTTTA  
 151 AAGGAAGACC GCGCGCGCAT CGTCGCCmAT ATGCGGCAGG CGGGTTTGAA  
 201 CCCCAGCCCC AAAACGGTCA AAGCCGTTTT TCGGAAACG GCAAAAGGCG  
 251 GTTTGGAAC TCCCCCGCG TTTTTCAGAA AACCGGAAGA CATAGAAACA  
 301 ATGTTCAAAG CGGTACACGG CTGGGAACAT GTGCAGCAGG CTTTGGACAA  
 351 ACACGAAGGG CTGCTATTC..

20 This corresponds to the amino acid sequence <SEQ ID 566; ORF138>:

1 MFRLQFRLFP PLRTAMHILL TALLKCLSL PLSCLHTLGN RLGHlafyLL  
 51 KEDRARIVAX MRQAGLNPD KTVKAVFAET AKGGLLAPA FFRKPEDiet  
 101 MFKAVHGEH VQALDKHEG LLF

Further work revealed the complete nucleotide sequence <SEQ ID 567>:

1 ATGTTTCGTT TACAATTCAG GCTGTTTCCC CCTTTGCGAA CCGCCATGCA  
 51 CATCCTGTTG ACCGCCCTGC TCAAATGCCT CTCCTGCTG CCGCTTTCCT  
 101 GTCTGCACAC GCTGGGAAAC CGGCTCGGAC ATCTGGCGTT TTACCTTTTA  
 151 AAGGAAGACC GCGCGCGCAT CGTCGCCAAT ATGCGGCAGG CGGGTTTGAA  
 201 CCCCAGCCCC AAAACGGTCA AAGCCGTTTT TCGGAAACG GCAAAAGGCG  
 251 GTTTGGAAC TCCCCCGCG TTTTTCAGAA AACCGGAAGA CATAGAAACA  
 301 ATGTTCAAAG CGGTACACGG CTGGGAACAT GTGCAGCAGG CTTTGGACAA  
 351 ACACGAAGGG CTGCTATTC TACGCGCGCA CATCGGCAGC TACGATTGG  
 401 GCGGACGCTA CATCAGCCAG CAGCTTCCGT TCCCGCTGAC CGCCATGTAC  
 451 AAACGCGCGA AAATCAAAGC GATAGACAAA ATCATGCAGG CCGGCAGGGT  
 501 TCGCGGCAAA GGAAAAACCG CGCCTACCAG CATAAAGGG GTCAAACAAA  
 551 TCATCAAAGC CCGTGCCTCG GCGGAAGCAA CCATCGTCCT GCCCGACCAC  
 601 GTCCCTCCC CTCAAGAAGG CGGGGAAGGC GTATGGGTGG ATTTCTTCGG  
 651 CAAACCTGCC TATACCATGA CGCTGGCGGC AAAATTGGCA CACGTCAAAG  
 701 GCGTGAAGAC CCTGTTTTTC TGCTGCGAAC GCCTGCCTGG CGGACAAGGT  
 751 TTCGATTGTC ACATCCGCCG CGTCCAAGGG GAATTGAACG GCGACAAAGC  
 801 CCATGATGCC GCCGTGTTC ACGCAATGC CGAATATTGG ATACGCCGTT  
 851 TTCCGACGCA GTATCTGTTT ATGTACAACC GCTACAAAAT GCCGTAA

This corresponds to the amino acid sequence <SEQ ID 568; ORF138-1>:

1 MFRLQFRLFP PLRTAMHILL TALLKCLSL PLSCLHTLGN RLGHlafyLL  
 51 KEDRARIVAN MRQAGLNPD KTVKAVFAET AKGGLLAPA FFRKPEDiet  
 101 MFKAVHGEH VQALDKHEG LLFITPHIGS YDLGGYISQ QLPFPLTAMY  
 151 KPPKIAIDK IMQAGVRGK GKTAPTSIQG VKQIIKALRS GEATIVLPDH  
 201 VPSPQEGGEG VWVDFGKPA YMTTLAAKLA HVKGVKTLFF CCERLPGGQG  
 251 FDLHIRPVQG ELNGDKAHDA AVFNRAEYV IRRFPTQYLF MYNRYKMP\*

50 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF138 shows 99.2% identity over a 123aa overlap with an ORF (ORF138a) from strain A of *N.*

*meningitidis*:

```

5      orf138.pep  MFR LQFRLFPPLRTAMHILLTALLKCLSLPLSCLHTLGNRLGHLAFYLLKEDRARIVAX
      orf138a     MFR LQFRLFPPLRTAMHILLTALLKCLSLPLSCLHTLGNRLGHLAFYLLKEDRARIVAN
10      orf138.pep  MRQAGLNPDPKTVKAVFAETAKGGLELAPAFFRKPEDIMTFKAVHGWEHVQQALDKHEG
      orf138a     MRQAGLNPDPKTVKAVFAETAKGGLELAPAFFRKPEDIMTFKAVHGWEHVQQALDKHEG
15      orf138.pep  LLF
      orf138a     LLFITPHIGSYDLGGYISQQLPFPLTAMYKPPKIKAIKIMQAGRVRGKGKTAPTTSIQG
20      orf138a     LLFITPHIGSYDLGGYISQQLPFPLTAMYKPPKIKAIKIMQAGRVRGKGKTAPTTSIQG

```

The complete length ORF138a nucleotide sequence <SEQ ID 569> is:

```

1  ATGTTTCGTT TACAATTCAG GCTGTTTCCC CCTTTGCGAA CCGCCATGCA
51 CATCCTGTTG ACCGCCCTGC TCAAATGCCT CTCCCTGCTG CCGCTTTCCT
101 GTCTGCACAC GCTGGGAAAC CGGCTCGGAC ATCTGGCGTT TTACCTTTTA
25 151 AAGGAAGACC GCGCGCGCAT CGTCGCCAAT ATGCGTCAGG CAGGCATGAA
201 TCCCGACCCC AAAACGGTCA AAGCCGTTT TGCGGAAACG GCAAAGGCG
251 GTTTGGAAC TGGCCCCGCG TTTTTCAGAA AACCGGAAGA CATAGAAACA
301 ATGTTCAAAG CGGTACACGG CTGGGAACAT GTGCAGCAGG CTTTGGACAA
351 ACACGAAGGG CTGCTATTCA TCACGCCGCA CATCGGCAGC TACGATTGG
401 GCGGACGCTA CATCAGCCAG CAGCTTCCGT TCCCGCTGAC CGCCATGTAC
451 AAACGCCCGA AATCAAAGC GATAGACAAA ATCATGCAGG CGGGCAGGGT
501 TCGCGGCAAA GGAAAAACCG CGCCTACCAG CATACAAGGG GTCAAACAAA
551 TCATCAAAGC CCTGCGTTCG GCGGAAGCAA CCATCGTCCT GCCCGACCAC
601 GTCCCTCCC CTCAAGAAGG CGGGGAAGGC GTATGGGTGG ATTTCTTCGG
35 651 CAAACCTGCC TATACCATGA CGCTGGCGGC AAAATTGGCA CACGTCAAAG
701 GCGTGAAAAC CCTGTTTTTC TGCTGCGAAC GCCTGCCTGG CGGACAAGGT
751 TTCGATTTGC ACATCCGCCG CGTCCAAGGG GAATTGAACG GCGACAAAGC
801 CCATGATGCC GCGGTGTTCA ACCGCAATGC CGAATATTGG ATACGCCGTT
851 TTCCGACGCA GTATCTGTTT ATGTACAACC GCTACAAAT GCCGTAA

```

40 This encodes a protein having amino acid sequence <SEQ ID 570>:

```

1  MFR LQFRLFP PLRTAMHILL TALLKCLSL PLSC LHTLGN RLGHLAFYLL
51  KEDRARIVAN MRQAGLNPD KTVKAVFAET AKGGLELAPA FFRKPEDIMT
101 MFKAVHGWEH VQALDKHEG LLFITPHIGS YDLGGYISQ QLPFPLTAMY
151 KPPKIKAIK IMQAGRVRGK GKTAPTTSIQG VKQIIKALRS GEATIVLPDH
45 201 VPSPQEGGEG VWVDFFGKPA YMTLAAKLA HVKGVKTLFF CCERLPGGQG
251 FDLHIREVQG ELNGDKAHDA AVFNRAEYW IRRFPTQYLF MYNRYKMP*

```

ORF138a and ORF138-1 show 99.7% identity over a 298aa overlap:

```

50 orf138a.pep  MFR LQFRLFPPLRTAMHILLTALLKCLSLPLSCLHTLGNRLGHLAFYLLKEDRARIVAN
      orf138-1   MFR LQFRLFPPLRTAMHILLTALLKCLSLPLSCLHTLGNRLGHLAFYLLKEDRARIVAN
      orf138a.pep  MRQAGMNPDPKTVKAVFAETAKGGLELAPAFFRKPEDIMTFKAVHGWEHVQQALDKHEG
      orf138-1     MRQAGLNPDPKTVKAVFAETAKGGLELAPAFFRKPEDIMTFKAVHGWEHVQQALDKHEG
55      orf138a.pep  LLFITPHIGSYDLGGYISQQLPFPLTAMYKPPKIKAIKIMQAGRVRGKGKTAPTTSIQG
      orf138-1     LLFITPHIGSYDLGGYISQQLPFPLTAMYKPPKIKAIKIMQAGRVRGKGKTAPTTSIQG
60      orf138a.pep  VKQIIKALRSGEATIVLPDHVPSPQEGGEGVWVDFFGKPAYMTLAAKLAHVKGVKTLFF

```

5

orf138-1	VKQIKALKRSEATIVLPDHVPS PQEGEGGVVWDFFGKPAYTMTLAAKLAHVKGVKTLFF
orf138a.pep	CCERLPGGGQGFDLHIRPVQGELNGDKAHDAAVFNRNAEYWIRRFTQYLFMYNRYKMP
orf138-1	CCERLPGGGQGFDLHIRPVQGELNGDKAHDAAVFNRNAEYWIRRFTQYLFMYNRYKMP

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF138 shows 94.3% identity over a 123aa overlap with a predicted ORF (ORF138ng) from

10 *N.gonorrhoeae*:

	orf138.pep	MFR LQFRLFPPLRTAMHILLTALLKCLSLPLSCLHTLGNRLGHLAFYLLKEDRARIVAX       :	60
	orf138ng	MFR LQFRLFPPLRTAMHILLTALLKCLSLLSLSC LHTLGNRLGHLAFYLLKEDRARIVAN       :	60
15	orf138.pep	MRQAGLNPD PKTVKAVFAETAKGGLE LAPAFFRKPED IETMF KAVHGWEHVQQALDKHEG       :	120
	orf138ng	MRQAGLNPD TQTVKAVFAETAKCGLE LAPAFFKKPED IETMF KAVHGWEHVQQALDKGEG       :	120
20	orf138.pep	LLF 	123
	orf138ng	LLFITPHIGSYDLGG RYISQQLPFH LTAMYKPPKIKAI DKIMQAGRVRGKGKTAPTGIQG       :	180

The complete length ORF138ng nucleotide sequence <SEQ ID 571> is:

	1	ATGTTTCGTT	TACAATTTCAG	GCTGTTTCCC	CCTTTGCGAA	CCGCCATGCA
25	51	CATCCTGTTG	ACCGCCCTGC	TCAAATGCCT	CTCCCTGCTG	TCGCTTTCTT
	101	GTCGTGCAC	GCTGGGAAAC	CGGCTCGGAC	ATCTGGCGTT	TTACCTTTTA
	151	AAGGAAGACC	GCGCGCGCAT	CGTCGCCAAT	ATGCGCGAGG	CTGGGTTTGA
	201	CCCCGACACG	CAGACGGTCA	AAGCCGTTTT	TGCGGAAACG	GCAAAATGCG
	251	GTTTGGAACT	TGCCCCCGCG	TTTTTCAAAA	AACCCGAAGA	CATCGAAACA
30	301	ATGTTTCAAG	CGGTACACGG	CTGGGAACAC	GTGACGAGG	CTTTGGACAA
	351	GGGCGAAGGG	CTGCTGTTCA	TCACGCCGCA	CATCGGCAGC	TACGATTTGG
	401	GCGGACGCTA	CATCAGCCAG	CAGCTTCCGT	TCCACTGCAC	CGCCATGTAC
	451	AAGCCGCCGA	AAATCAAAGC	GATAGACAAA	ATCATCAGG	CGGGCAGGGT
	501	GCGCGGCAAA	GCGAAAAACcg	cgcaccgcg	atACAAGGG	GCTCAACAAA
35	551	tcatcaAGGC	CCTGCGCGCG	GGCGAGGCAA	CCAtcATCCT	GCCCCAGCCA
	601	GTCCCTTCTC	CGCAGGAagg	cggCGGCGTG	TGGGCGGATT	TTTTTCGGCA
	651	ACCTTGCAAT	acCATGACAC	TGGCGGCAAA	ATTGGCACAC	GCTCAAAGCG
	701	TGAAAACCTT	GTTTTTCTGC	TGCGAACGCC	TGCCCCGACG	ACAAGGCTTC
	751	GTGTTGCACA	TCCGCCCCGT	CCAAGGGGAA	TTGAACGGCA	ACAAAGCCCA
40	801	CGATGCCGCC	GTGTTCAACC	GCAATACCGA	ATATTGGATA	CGCCGTTTTT
	851	GACACGAGTA	TCTGTTTATG	TACAACCGCT	ATAAAACGCC	GTAA

This encodes a protein having amino acid sequence <SEQ ID 572>:

```

1 MFR LQFRLFP PLRTAMHILL TALLKCLSL SLSCLHTLGN RLGLHAFYLL
51 KEDRARIVAN MRQAGLNPD TQVKAVFAET AKCGLELAPA FFFKKPEDIET
101 MFKAVHGEWH VQALDKGEG LLFITPHIGS YDLGGYISQ QLPFHLTAMY
45 151 KPPKIKAIK IMQAGRVRGK GKTAPTGIQG VKQIIKALRA GEATIILPDH
201 VSPSQEGGVG WADFFGKPAY TMTLAACKLA VKGVTLFFCC CERLPDQGGF
251 VLHIRPVQGE LNGNKAHDA VFNRNTEYWI RRFPTQYLFM NRYNKP*

```

ORF138ng and ORF138-1 show 94.3% identity over 299aa overlap:

[illegible]

```

orfl38-1.pep  VKQIIKALRSGEATIVLPDHVPSPQEGGEGVWVDFGKPAYTMTLAAKLAHVKGVKTLEFF
|||||||:|||||:||||||| |||:|||||||
orfl38ng      VKQIIKALRAGEATIILPDHVPSPQEGG-GVWADFFGKPAYTMTLAAKLAHVKGVKTLEFF

5  orfl38-1.pep  CCEERLPGGQGFDLHIRPVQGELNGDKAHDAAVFNRNAEYWIRRFPTQYLFMYNRYKMP
||||| ||| |||||||||:|||||||:|||||||
orfl38ng      CCEERLPGGQGFVLHIRPVQGELNGNKAHDAAVFNRNTEYWIRRFPTQYLFMYNRYKTP

```

In addition, ORF138ng is homologous to htrB protein from *Pseudomonas fluorescens*:

```

gnl|PID|e334283 (Y14568) htrB [Pseudomonas fluorescens] Length = 253
10  Score = 80.8 bits (196), Expect = 9e-15
    Identities = 49/151 (32%), Positives = 79/151 (51%), Gaps = 6/151 (3%)

Query: 101 MFKAVHGWVHVQQALDKGEGLLFITPHIGSYD-LGGRYISQQLPFHLTAMYKPPKIKAI 159
      + + V G E +++AL G+G++ IT H+G+++ L Y SQ P Y+PPK+KA+D
15  Sbjct: 94 LVREVEGLEVLKEALASGKGVVITSHLGNWEVLNHFYCSQCKPI---IFYRPPKLKAVD 150

Query: 160 KIMQAGRVRGKGTAPTGIQGVKQIIKALRAGEATIILPDHVPSPQEGGGVWADFFGKPA 219
      +++++ RV+ K A + +G+ +IK +R G I D P P E G++ FF A
20  Sbjct: 151 ELLRKQRVQLGNKVAASTKEGILSVIKEVRKGGQVGIPAD--PEPAESAGIFVFFATQA 208

Query: 220 YTMTLAAKLAHVKGVKTLEFFCCERLPGDGGF 250
      T + +F RLPDG G+
25  Sbjct: 209 LTSKFVPNMLAGGKAVGVFLHALRLPDGSGY 239

```

Based on this analysis, including the presence of a putative transmembrane domain in the  
 25 gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and  
 their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF138-1 (57kDa) was cloned in the pGex vectors and expressed in *E.coli*, as described above.  
 The products of protein expression and purification were analyzed by SDS-PAGE. Figure 14A  
 shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein  
 30 was used to immunise mice, whose sera were used for ELISA (positive result) and FACS analysis  
 (Figure 14B). These experiments confirm that ORF138-1 is a surface-exposed protein, and that it  
 is a useful immunogen.

### Example 69

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 573>:

```

35      1 ..CGCTGGTCCG CCGGCGAATC GTGGCGTGTG TTAATGGAAA GTGAAACGTG
      51 GCATGCGGTG TGGAATACTT TGCGCTTCTC GGCGGCGGCG GTGTATGCGG
      101 CAGCGGTTTT GGGTGTGGTG TATGCGGCGC CGGCGCGGCG GTCGGCGTGG
      151 ATGCGCGGGC TGATGTTTGA GCCGTTTATG GTGTCGCCGG TTTGTGTTTC
      201 GGCGGGCGTG CTGCTGCTTT ATCCGCAGTG GACGGCTTCG TTGCCGTTGC
      251 TGCTGGCGAT GTATGCGCTG CTGGCGTATC CGTTTGTGGC AAAAGATGTT
40      301 TTATCAGCCT GGGATGCACT GCCGCCGAT TACGGCAGGG CGGCGGCGGG
      351 TTTGGGTGCA AACGGCTTTC AGACGGCATG CCGCATCACG TCCCCCTCT
      401 TGAAACCGGC GTTGCAGGCG GGTCTGACTT TGCGGCGGCG AACCTGCGTG
      451 GGCGAATTG CCGCGACATT GTTCTGTGCG CGTCCGGAAT GGCAGACGCT
45      501 GACGACTTTG ATTTATGCCT ATTTGGGACG CGCGGGTGAG GATAATTACG
      551 CGCGGGCGAT GGTGCTG..

```

This corresponds to the amino acid sequence <SEQ ID 574; ORF139>:

```

      1 ..AWSAGESWRV LMESETWHAV WNTLRFSAVA VYAAAVLGVV YAAPARRSAW
      51 MRGLMFXPFM VSPVCVSAGV LLLYPQWTAS LPLLLAMYL LAYPFVAKDV

```



101 LSAWDALPPD YGRAAGLGA NGFQTACRIT FPLLKPALRR GLTLAAATCV  
151 GEFAATLFLS RPEWQTLTTL IYAYLGRAGE DNYARAMVL..

Further work revealed the complete nucleotide sequence <SEQ ID 575>:

5 1 ATGGATGGAC GCGGTTGGGT GGTATGGGGT GCTTTTGCCC TGCTGCCTTC  
51 GGCTTTTTTG GCGGTAATGG TCGTTGCGCC TTTGTGGGCG GTGGCGGCGT  
101 ATGACGGTTT GCGGTGGCGC GCGGTGCTGT CGGATGCCTA TATGCTCAAA  
151 CGTTTGGCGT GGACGGTATT TCAGGCAGCG GCAACCTGTG TGCTGGTGCT  
201 GCCTTTGGGC GTGCCTGTCG CGTGGGTGCT GGCGCGGCTG GCGTTTCCGG  
251 GGCGGGCTTT GGTGCTGCGC CTGCTGATGC TGCTTTTGTGATGCCACG  
10 301 TTGGTGGCGG GCGTGGGCGT GCTGGCCCTG TTCGGGGCGG ACGGGCTGTT  
351 GTGGCGGCGC AGGCAGGATA CGCCGTATCT GTTGTGTGAC GGCAATGTGT  
401 TTTTCAACCT TCCTGTGTTG GTCAGGGCGG CGTATCAGGG GTTGTGCAA  
451 GTGCCTGCGG CACGGCTTCA GACGGCACGG ACGTTGGGCG CGGGGCGGTG  
501 GCGGCGGTTT TGGGACATTG AAATGCCCGT TTTGCGCCCG TGGCTTGCCG  
15 551 GCGGCGGTG CTGTGCTTT CTGTATTGTT TTTCCGGGTT CGGGCTGGCG  
601 CTGCTGCTGG GCGGCAGCCG TTATGCCACG GTCGAAGTGG AAATTTACCA  
651 GTTGGTCATG TTCGAACCTG ATATGGCGGT TGCTTCGGTG CTGGTGTGGC  
701 TGGTGTGGG GGTAAACGGG GCGCAGGGT TGCTGTATGC GTGGTTCGCG  
751 AGGCGCGCGG TTTCCGATAA GCGGTTTCC CTTGTGATGC CGTCGCCGCC  
20 801 GCAGTCGGTC GGGGAATATG TGCTGCTGCG GTTGTGCGCG GCGGTGTGTG  
851 CTGTGTGCTG CTTGTTTCC TTTGTGGCAA TTGTGTGAA AGCGTGGTCG  
901 GCGGCGAAT CGTGGCGTGT GTTAATGGAA AGTGAAACGT GCGAGGCGGT  
951 GTGAATACT TTGCGCTTCT GCGGCGCGG GGTGTATGCG GCGGCGGTTT  
1001 TGGGTGTGGT GTATGCGGCG GCGGCGGCG GGTGCGCGTG GATGCGCGGG  
25 1051 CTGATGTTTT TGCCGTTTAT GGTGTCGCG GTTGTGTGTT CGGCGGCGGT  
1101 GCTGCTGCTT TATCCGCACT GGACGGCTTC GTTGCCTTG CTGCTGGCGA  
1151 TGTATGCGCT GCTGGCGTAT CCGTTTGTGG CAAAAGATGT TTATCAGCC  
1201 TGGGATGCAC TGCCGCCGGA TTACGGCAGG GCGGCGGCG GTTGGGTGTC  
1251 AAACGGCTTT CAGACGGCAT GCCGCATCAC GTTCCCCTC TTGAAACCGG  
30 1301 CGTTGCGGCG CGGTCTGACT TTGGCGGCG CAACCTGCGT GGGCGAATTT  
1351 GCGGCGACAT TGTCTCTGTC GCGTCCGGAA TGGCAGACGC TGACGACTTT  
1401 GATTTATGCC TATTTGGGAC GCGCGGCTGA GGATAATTAC GCGGCGGCGA  
1451 TGGTCTGAC ATTCTGTTG GCGGCGTTCG CGCTGGGTAT TTTCTGCTG  
1501 TTGGACGGCG GCGAAGGCG AAAACAGACG GAAACGTTAT AA

35 This corresponds to the amino acid sequence <SEQ ID 576; ORF139-1>:

1 MDGRRVWVG AFALLPSAFL AVMVVAFLWA VAAVDGLAWR AVLSDAYMLK  
51 RLAWTVFQAA ATCVLVPLG VPVAVLRL AFPGRALVLR LLMLPFVMT  
101 LVAGVGVLAL FGADGLLWRG RQDTPYLLLY GNVFFNLPVL VRAAYQGFVQ  
151 VPAARLQATAR TLGAGAWRRF WDIEMPVLRP WLAGGVCLVF LYCFSGFGLA  
40 201 LLLGGSRYAT VEVEIYQLVM FELDMAVASV LVWLVLGVTA AAGLLYAWFG  
251 RRAVSDKAVS PVMPSPPQSV GEYVLLAFAA AVLSVCCLFP LLAIVVKAWS  
301 AGESWRVLM SETWQAVWNT LRFSAAVYA AAVLGVVYAA AARRSAWMRG  
351 LMFLPFMVSP VCVSAGVLLL YPQWTASLPL LLAMYALLAY PFVAKDVLSA  
401 WDALPPDYGR AAAGLGANGF QTACRITFPL LKPALRRGLT LAAATCVGEF  
45 451 AATLFLSRPE WQTLTLIYA YLGRAGEDNY ARAMVLTLLL AAFALGIFLL  
501 LDGEGGKQT ETL\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF139 shows 94.7% identity over a 189aa overlap with an ORF (ORF139a) from strain A of *N.*

50 *meningitidis*:

orfl39.pep 10 20 30  
AWSAGESWRVLMESETHAVWNTLRFSAAA  
orfl39a QSVGEYVLLAFAAAVXSVCCLEFXLLAIVVKAWSAGESWRVLMESETHAVWNTXRFSA  
55 270 280 290 300 310 320  
orfl39.pep 40 50 60 70 80 90  
VYAAAVLGVVYAAPARRSAWMRGLMFXPFMVSPVCVSAGVLLLYPQWTASLPLLLAMYAL  
|||||

orf139a VYAAAVLGVVYAAAARRSAWMRGLMFLPFMVSPVCVSAGVLLLPQWTASLPLLLAMYAL  
 330 340 350 360 370 380

5 orf139.pep LAYPFVAKDVLSAWDALPPDYGRAAGLGANGFQTACRITFPLLKPALRRGLTLAAATCV  
 |||||  
 orf139a LAYPFVAKDVLSAXDALPPDYGRAAGLGANGFQTACRITFPLLKPALRRGLTLAAATCV  
 390 400 410 420 430 440

10 orf139.pep GEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNYARAMVL  
 |||||  
 orf139a GEFAATLFXSRXEWQTLTTLIYAYXGRAGXDNYARAMVLTLLLAALGXFLLDGEGG  
 450 460 470 480 490 500

15 The complete length ORF139a nucleotide sequence <SEQ ID 577> is:

1 ATGGATGGAC GCGGTTGGGC GGTATGGGGT GCTTTTGCCC TGCTGCCTTC  
 51 GGCTTTTTTG GCGGCAATGG TCGTTGCGCC TTTGTGGGCG GTGGCGGCGT  
 101 ATGACGGTTT GCGGTGGCGC GCGGTGCTGT CGGATGCCTA TATGCTCAAA  
 151 CGTTTGGCGT GGACGGTATT TCAGGCAGCG GCAACCTGTG TGCTGGTGCT  
 201 GCCTTTGGGC GTGCCTGTCG CGTGGGTGCT GCGCGGGCTG CGCTTCCGG  
 251 GCGGGGCTTT GGTGCTGCGC CTGCTGATGC TGCTTTTGT GATGCCACG  
 301 TTGGTGGCGG GCGTGGGCGT GCTGGCTCTG TTCGGGGCGG ACGGCCTGTN  
 351 GTGGCGCGGC TGGCAGGATA CGCGGTATCT GTTGTGTAC GGCAATGTGT  
 401 TTTTNNACCT TCCTGTGTTG GTCAGGGCGG CATATCAGGG GTTGTGCAA  
 251 GTGCTGCGG CACGGCTTCA GACGGCACNG ACATTGGGCG CGGGGGCGTG  
 501 GCGGCGGTTT TGGGACATTG AAATGCCCGT TTTGCGCCCG TGGCTTGCCG  
 551 GCGGCGGTG CTTGTCTTC CTGATTGTT TTTGCGGGTT CGGGCTGGCA  
 601 TTGCTGCTGG GCGGCAGCCG TTATGCCACG GTCGAAGTGG AAATTTACCA  
 651 GTTGGTCATG TTGCAACTCG ATATGGCGGT TGCTTCGGTG CTNGTGTGC  
 301 TGGTGTNGG GGTAAACNGC GCGGCAGGCT TGCTGTATGC GTGTTCCGGC  
 751 AGCGCGCGG TTTCCGATAA GCGNGTTTCC CCTGTGATGC CGTCGCGGCC  
 801 GCAGTCGGTC GGGGAATATG TGCTNCTGGC GTTTCGGCGC GCGGTGTNGT  
 851 CTGTGTGCTG CCTGTTTCNT TTGTGGCAA TTGTGTGAA AGCGTGGTGC  
 901 GCGGCGCAAT CGTGGCGTGT GTTAATGAA AGTGAAACGT GGCAGGCGGT  
 351 GTGGAATACT NTGCGCTTCT CGGCGGCGGC GGTGTATGCG GCGGCGGTTT  
 1001 TGGGTGTGGT GTATGCGGCG GCGGCGCGGC GGTGCGCGTG GATGCGCGGG  
 1051 CTGATGTTTT TGCCGTTTAT GGTGTCGCGC GTTGTGTTT CGGCGGGCGT  
 1101 GCTGCTGCTT NATCCGAGT GGACGGCTTC GTTCCGCGTG CTGCTGGCGA  
 1151 TGTATGCGCT GCTGGCGTAT CCGTTTGTGG CAAAAGATGT TTTATCAGCC  
 401 TNGATGCAC TGCGCCGGA TTACGGCAGG GCGGCGGCGG GTTGGGTGC  
 1201 AAACGGCTTT CAGACGGCAT GCCGCATCAC GTTCCCCCTC TTGAAACCGG  
 1251 CGTTGCGGCG CGGTCTGACT TTGGCGGCGG CAACCTGCGT GGGCGAATTT  
 1301 GCGGCAACCT TGTTCNTGTC GCGTCNCGAG TGGCAGACGC TGACGACTTT  
 1351 GATTTATGCC TATNTGGGAC GCGCGGGTGA NGATAATTAC GCGCGGGCGA  
 451 TGGTGCTGAC ATTGCTGTTG GCGGCGTTCG CGCTGGGTAT NTTCTGCTG  
 1401 TTTGACGGCG GCGAAGGCGG AAAACGGACG GAAACGTTAT AA

This encodes a protein having amino acid sequence <SEQ ID 578>:

1 MDGRRWAVWG AFALLPSAFL AAMVVAPLWA VAAYDGLAWR AVLSDAYMLK  
 51 RLAWTVFQAA ATCVLVPLG VPVAVVLARL AFPGRALVLR LLMLPFVMPT  
 101 LVAGVGVLAL FGADGLXWRG WQDTPYLLLY GNVFFXLPVL VRAAYQGFQ  
 151 VPAARLQTAX TLGAGAWRRF WDIEMPVLRP WLAGGVCLVF LYCFSGFGLA  
 201 LLLGGSRYAT VEVEIYQLVM FELDMAVASV LVWLXGVTA AAGLLYAWFG  
 251 RRAVSDKAWS PVMPSPPQSV GEYVLLAFAA AVXSVCCLFX LLAIVVKAWS  
 301 AGESWRVLME SETWQAVWNT XRFSAAYYA AAVLGVVYAA AARRSAWMRG  
 55 351 LMFLPFMVSP VCVSAGVLLL XPQWTASLPL LLAMYALLAY PFVAKDVLSA  
 401 XDALPPDYGR AAAGLGANGF QTACRITFPL LKPALRRGLT LAAATCVGEF  
 451 AATLFXSRXE WQTLTTLIYA YXGRAGXDNY ARAMVLTLLL AAFALGXFLL  
 501 LDGEGGGKRT ETL\*

ORF139a and ORF139-1 show 96.5% homology over a 514aa overlap:

60 orf139a.pep MDGRRWAVWGAFALLPSAFLAAMVVAPLWAVAAAYDGLAWRAVLSDAYMLKRLAWTVFQAA  
 |||||:|||||  
 orf139-1 MDGRRWVWGAFAALLPSAFLAMVVAPLWAVAAAYDGLAWRAVLSDAYMLKRLAWTVFQAA

	orf139a.pep	ATCVLVLPPLGVPVAVVLARLAFPGRALVLRLLMLPFVMPTLVAGVGVLAALFGADGLXWRG
	orf139-1	ATCVLVLPPLGVPVAVVLARLAFPGRALVLRLLMLPFVMPTLVAGVGVLAALFGADGLLWRG
5	orf139a.pep	WQDTPYLLLYGNVFFXLPVLVRAAYQGFVQVPAARLQTAXTLGAGAWRRFWDIEMPVLRP
	orf139-1	RQDTPYLLLYGNVFFNLPVLVRAAYQGFVQVPAARLQTARTLGAGAWRRFWDIEMPVLRP
10	orf139a.pep	WLAGGVCLVFLYCFSGFGLALLLGGSSRYATVEVEIYQLVMFELDMAVASVLVWLVLXGVTA
	orf139-1	WLAGGVCLVFLYCFSGFGLALLLGGSSRYATVEVEIYQLVMFELDMAVASVLVWLVLGVTA
15	orf139a.pep	AAGLLYAWFGRRRAVSDKAVSPVMPSPQSVGEYVLLAFAAAVXSVCCFLXLLAIVVKAWS
	orf139-1	AAGLLYAWFGRRRAVSDKAVSPVMPSPQSVGEYVLLAFAAAVLSVCCFLPLLAIVVKAWS
20	orf139a.pep	AGESWRVLMESSETWQAVWNTXRFSAAAVYAAAVLGVVYAAAARRSAWMRGLMFLPFMVSP
	orf139-1	AGESWRVLMESSETWQAVWNTLRFSAAVYAAAVLGVVYAAAARRSAWMRGLMFLPFMVSP
25	orf139a.pep	VCVSAGVLLLLXPQTASLPLLLAMYALLAYPFVAKDVLSAXDALPPDYGRAAAGLGANGF
	orf139-1	VCVSAGVLLLLYPQTASLPLLLAMYALLAYPFVAKDVLSAWDALPPDYGRAAAGLGANGF
30	orf139a.pep	QTACRITFLLKPALRRGLTLAAATCVGEFAATLFXSRKEWQTLTTLIYAYXGRAGXDNY
	orf139-1	QTACRITFLLKPALRRGLTLAAATCVGEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNY
	orf139a.pep	ARAMVLTLLLAALFALGXFLLLDGGEGGKRTETLX
	orf139-1	ARAMVLTLLLAALFALGIFLLLDGGEGGKQTETLX

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF139 shows 95.2% identity over a 189aa overlap with a predicted ORF (ORF139ng) from *N.gonorrhoeae*:

35	orf139.pep	AWSAGESWRVLMESSETWHA VWNTRFSAAA	30
	orf139ng	QSVGEYVLLAFSVAVLSVCCFLPLSAIVVKAWSAGESRRVLMESSETWQAVWNTLRFSAAA	327
40	orf139.pep	VYAAAVLGVVYAAPARRSAWMRGLMFXPFMVSPVCVSAGVLLLYPQWTASLPLLLAMYAL	90
	orf139ng	VFAAAVLGVVYAAAARRLVWMRGLVLFPEFMVSPVCVSAGVLLLYPGWTASLPLLLAMYAL	387
45	orf139.pep	LAYPFVAKDVLSAWDALPPDYGRAAAGLGANGFQTACRITFLLKPALRRGLTLAAATCV	150
	orf139ng	LAYPFVAKDVLSAWDALPPDYGRAAAGLGANGFQTACRITFLLKPALRRGLTLAAATCV	447
	orf139.pep	GEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNYARAMVL	189
50	orf139ng	GEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNYARAMVLTLLLSAFAVCIFLLLDNGEGG	507

The complete length ORF139ng nucleotide sequence <SEQ ID 579> is predicted to encode a protein having amino acid sequence <SEQ ID 580>:

55	1	MDGRCAVVRG	AFSLLPSAFL	AVMVVAPLWA	VAAYDGLAWR	AVLSDAYMLK
	51	RLAWTVFQAA	ATCVLVLPPLG	VPVAVVLARL	AFPGRALVLR	LLMLPFVMPT
	101	LVAGVGVLAAL	FGADGLLWRG	RQDTPYLLLY	GNVFFNLPVL	VRAAYQGFQAQ
	151	VPAARLOTAR	TLGAGAWRPF	WDIEMPVLRP	WLAGGVCLVF	LYCFSGFGLA
	201	LLLGGSSRYAT	VEVEIYQLVM	FELDMAGASA	LVWLVLGVTA	AAGLLYAWFG
	251	RRAVSDKAVS	PVMPSPQSV	GEYVLLAFSV	AVLSVCCFLP	LSAIVVKAWS
	301	AGESRRVLMES	SETWQAVWNT	LRFSAAAVFA	AAVLGVVYAA	AARRLVWMRG
60	351	LVFLPFMVSP	VCVSAGVLLL	YPGWTASLPL	LLAMYALLAY	PFVAKDVLSA
	401	WDALPPDYGR	AAAGLGANGF	QTACRITFPL	LKPALRRGLT	LAAATCVGEF
	451	AATLFLSRPE	WQTLTTLIYA	YLGRAGEDNY	ARAMVLTLLL	SAFAVCIFLL
	501	LDNGEGGKRT	ETL*			

Further work revealed a variant gonococcal DNA sequence <SEQ ID 581>:

	1	ATGGATGGAC	GGTGTGGGC	GGTACGGGGT	GCTTTTTTCC	TGCTGCCTTC
	51	GGCTTTTTTG	GCGGTAATGG	TCGTTGCGCC	TTTGTGGGCG	GTGGCGGCGT
5	101	ATGACGGTTT	GGCGTGGCGC	GCGGTGCTGT	CGGATGCCTA	TATGCTCAAA
	151	CGTTTGGCGT	GGACGCTGTT	TCAGGCGGCG	GCAACCTGTG	TGCTGGTGCT
	201	GCCTTTGGGC	GTGCCGTGTCG	CGTGGGTGCT	GGCGCGGCTG	GCGTTCGCCG
	251	GGCGGGCCTT	GGTGCTGCGC	CTGCTGATGC	TGCCGTTTGT	GATGCCACAG
	301	CTGGTGGCGT	GCGTGGGCGT	GCTGGCTCTG	TTTGGGGCGG	ACCGGCTGTT
10	351	GTGGCGCGCG	GCGGAGGATA	CGCCGTATCT	GTCGTTGTAC	GCGAATGTGT
	401	TTTTCAACCT	GCCCCGTGTTG	GTCAGGCGCG	CGTATCAGGG	GTTTGCTCAA
	451	GTGCCGTGCG	CACGGCTTCA	GACGGCACGG	ACGTTTGGGCG	CGGGGGCGTG
	501	GCGGCGGTTT	TGGGACATTG	AAATGCCCGT	TTTGCGCCCG	TGGCTTGCCG
	551	GCGGCGGTGT	CCTTGCTTTC	CTGTATTGTT	TTTCGGGGTT	GCGGCTGGCA
15	601	TTGCTGTTGG	GCGGCAGCCG	TTATGCCACG	GTCGAAGTGG	AAATTTACCA
	651	GTTGGTTATG	TTCGAACTCG	ATATGGCGGG	GGCTTCGGCG	CTGGTGTGGC
	701	TGGTGTTGGG	GGTAACGGCG	GCGGCAAGGT	TGCTGTATGC	GTGGTTCCGG
	751	AGGCGCGCGG	TTCGGAATAA	GGCGGTTTCC	CCCGTGATGC	CGTCGCGCGC
	801	GCAATCGGTG	GGGGAATATG	TATTGCTGGC	ATTTTCGGTG	GCGGTGTTGT
20	851	CCGTGTGCTG	CCTGTTTCTT	TTGTGCGCAA	TTGTTGTGAA	AGCGTGGTGC
	901	GCCGCGCAAT	GCGGCGGTGT	GTTAATGAACT	GTTGTAACGT	GCGAGGAGT
	951	TGGGAATACT	ttGCGCTTTT	CGGCGGCGGC	GGTGTTTGCG	GCGGCGGTTT
	1001	TGGGTGTGGT	GTATGCGGCG	GCGGCGCGGC	GGCTGGTGTG	GATGCGCGGA
	1051	CTGGTGTGTTT	TACCGTTTAT	GGTGTCGCCG	GTTTGTGTTT	CGGCGGGCGT
25	1101	GCTGCTGCTT	TATCCGGGGT	GGACGGCTTC	GTTACCGCTG	CTGCTGGCGA
	1151	TGTATGCGCT	GCTGGCGTAT	CCGTTTGTGG	CAAAAGATGT	TTTATCGGCC
	1201	TGGGATCGAC	TGCCGCGCGA	TTACGGCAGG	GCGGCGGCAG	GTTTGGGCGC
	1251	AAACGGCTTT	CAGACGGCAT	GCCGTATCAC	GTTCCCCCTC	TTGAAACCGG
	1301	CGTTGCGGCG	GGTCTGACT	TTGGCGGCGG	CGACGTGTGT	GGCGGTAATT
30	1351	GCGGCAACCT	TGTTCTGTGC	GCGTCCGGAA	TGGCAGACGT	TGACGACTTT
	1401	GATTTATGCC	TATTTGGGCG	GTGCGGGTGA	GGACAATTAT	GTCGCGGCAA
	1451	TGGTGTTGAC	ATTGCTGTTG	TCGGCATTTG	CGGTGTGCAT	TTTCTGCTG
	1501	TGGACAAACG	GCGAAGCGGq	aaaACGGACG	GAAACGTTAT	AA

This corresponds to the amino acid sequence <SEQ ID 582; ORF139ng-1>:

35	1	MDGRCWAVRG	<u>AFSLLPSAFL</u>	<u>AVMVVAPLWA</u>	VAAYDGLAWR	AVLSDAYMLK
	51	RLAWTVFQAA	<u>ATCVLVLP LG</u>	<u>VFVAWVLARL</u>	AFPGRALVLR	LMLPFVMPT
	101	<u>LVAGVGLVAL</u>	<u>FGADGLLWRG</u>	<u>RQDTPYLLLY</u>	GNVFFNLPLV	VRAAYQGFA
	151	VPAARLQTAR	TLGAGAWRRF	WDIEMPVLRP	VLAGGVCLVF	LYCFSGFGLA
40	201	<u>LLLGGSRyat</u>	VEVEIYQLVM	FELDMAGASA	<u>LVWLVLGVTA</u>	<u>AAGLLYAWFG</u>
	251	RRAVSDKAVS	PVMPSPQQSV	GEYVLLAFSV	<u>AVLSVCCFLP</u>	LSAIVVKWMS
	301	AGESRRVLME	SETWQAVWNT	LRFSAAAVFA	<u>AAVLGVVYAA</u>	AARRLVKWRG
	351	<u>LVFLPFMVSP</u>	<u>VCVSAGVLLL</u>	<u>YPGWtASLPL</u>	<u>LLAMYALLAY</u>	<u>PFVAKDVLSA</u>
	401	WDALPPDYGR	AAAGLGANGF	QTACRITFPL	LKPALRRGLT	LAATTCVGEF
	451	AATLFLSRPE	<u>WQTLTTLIYA</u>	YLGRAGEDNY	<u>ARAMVLTLLL</u>	<u>SAFAVCIPLL</u>
	501	LDNEGGKRT	ETL*			

45 ORF139ng-1 and ORF139-1 show 95.9% identity over 513aa overlap:

[illegible]

orf139ng AGESRRVLMES~~ETWQAVWNTLRFSAAAVFAAAVLGVVYAAAARRLVWMRGLVFLPFMVSP~~  
 orf139 AGESWRVLMES~~ETWQAVWNTLRFSAAAVFAAAVLGVVYAAAARRSAWMRGLMFLPFMVSP~~  
 5 orf139ng VCVSAGVLLLLPGWTASLPLLLAMYALLAYPFVAKDVLSAWDALPPDYGRAAAGLGANGF  
 orf139-1 VCVSAGVLLLLPQWTASLPLLLAMYALLAYPFVAKDVLSAWDALPPDYGRAAAGLGANGF  
 10 orf139ng QTACRITFPLLPALRRGLTAAATCVGEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNY  
 orf139-1 QTACRITFPLLPALRRGLTAAATCVGEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNY  
 15 orf139ng ARAMVLTLLLSAFAVCIFLLLDNGEGGKRTETL  
 orf139-1 ARAMVLTLLLA~~AFALGIFLLLDGEGGKQTETL~~

Based on the presence of a predicted binding-protein-dependent transport systems inner membrane component signature (underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 70

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 583>:

1 ATGGACGGCT GGACACAGAC GCTGTCCGCG CAAACCCTGT TGGGCATTTCT  
 51 GGCGGCGGCA ATCATCCTCA TTCTGATTTT AATCGTCAGA TTCCGCATCC  
 101 ACGCGCTGCT GACACTGGTC ATCGTCAGCC TGCTGACGGC TTTGGCAACC  
 151 GGTTTGCCCA CAGGCAGCAT TGTCAAAGAC ATACTGGTCA AAAACTTCGG  
 201 CGGCACGCTC GGCGGCGTGG CGCTTCTGGT CGGCCTGGGC GCGATGCTCG  
 251 AACGTTTGGT C...

This corresponds to the amino acid sequence <SEQ ID 584; ORF140>:

1 MDGWTQTL~~SA~~ QTL~~L~~GISAAA IILILILIVR FRIHALLTLV IVSLLTALAT  
 51 GLPTGSIVKD ILVKNFGGTL GGVALLVGLG AMLERLV..

Further work revealed the complete nucleotide sequence <SEQ ID 585>:

1 ATGGACGGCT GGACACAGAC GCTGTCCGCG CAAACCCTGT TGGGCATTTCT  
 51 GGCGGCGGCA ATCATCCTCA TTCTGATTTT AATCGTCAAA TTCCGCATCC  
 101 ACGCGCTGCT GACACTGGTC ATCGTCAGCC TGCTGACGGC TTTGGCAACC  
 151 GGTTTGCCCA CAGGCAGCAT TGTCAAAGAC ATACTGGTCA AAAACTTCGG  
 201 CGGCACGCTC GGCGGCGTGG CGCTTCTGGT CGGCCTGGGC GCGATGCTCG  
 251 GACGTTTGGT CGAAACATCC GGCGGCGCAC AGTCGCTGGC GGACGCGCTG  
 301 ATCCGGATGT TCGGCGAAAA ACGGCGACCG TTCGCGCTGG GCGTTGCCTC  
 351 GCTGATTTTC GGCTTCCCGA TTTTCTTCGA TGCCGGACTA ATCGTCATGC  
 401 TGCCCATCGT GTTCGCCACC GCACGGCGCA TGAACAGGA CGTACTGCC  
 451 TTCGCGCTTG CCTCCATCGG CGCATTTTCC GTCATGCACG TCTTCCTGCC  
 501 GCCCCATCCG GGCCCGATTG CCGCTTCCGA ATTTTACGGC GCGAACATCG  
 551 GCCAAGTTTT GATTTTGGGT CTGCCGACCG CCTTCATCAC ATGGTATTTT  
 601 AGCGGCTATA TGCTCGGCAA AGTGTGGGG CGCACCATCC ATGTTCCTCG  
 651 TCCCGAAGTG CTCAGCGGCG GCACGCAAGA CAACGACCTG CCGAAAGAAC  
 701 CTGCCAAGC AGGAACGGTC GTGCGCATCA TGCTGATTCC CATGCTGCTG  
 751 ATTTTCCTGA ATACCGCGT ATCGGCCCTC ATCAGCGAAA AACTCGTAAG  
 801 TCGGACGAA ACCTGGGTTC AGACGGCAA AATAATCGGT TCGACACCGA  
 851 TCGCCCTTCT GATTTCCGTA TTGGTCGCAC TGTTTGCTT GGGACGCAA  
 901 CGCGGCGAAA GCGGCAGCGC GTTGGAAAA ACCGTGGACG GCGCACTCGC  
 951 CCCCCTCTGT TCCGTGATTC TGATTACCGG CGCGGCGGT ATGTTCGGCG  
 1001 GCGTTTTCG CGCTTCCGCG ATCGGCAAGG CACTCGCCGA CAGCATGGCG  
 1051 GATTTGGGCA TCCCGTCTT TTTGGGCTGT TTCCTTGTCG CCTTGGCACT  
 1101 GCGTATCGCG CAAGGTTTCG CAACCGTCGC CCTGACCACC GCCGCGCGCG  
 1151 TGATGGCTCC TGCCGTTGCC GCCGCGGCT TTACCGACTG GCAGCTCGCC

1201 TGTATCGTAT TGGCAACGGC GGCAGGTTTC GTCCGTTGCA GCCACTTCAA  
 1251 CGACTCCGGC TTCTGGCTGG TCGGCCGTCT CTGGACATG GACGTACCGA  
 1301 CCACGCTGAA AACCTGGACG GTCAACCAAA CCTCATCGC ACTCATCGGC  
 1351 TTTGCCTTGT CCGCACTGCT GTTCGCCATC GTCTGA

5 This corresponds to the amino acid sequence <SEQ ID 586; ORF140-1>:

1 MDGWTQTLA QTLGLISAAA IILILILIVK FRIHALLTV IVSLLTALAT  
 51 GLPTGSIVND ILVKNFGGTL GGVALLVGLG AMLGRLVETS GGAQSLADAL  
 101 IRMGKRAP FALGVASLIF GFPIFFDAGL IVMLPIVFAT ARRMKQDVLV  
 151 FALASIGAFS VMHVLPHP GPIAASEFYG ANIGQVLILG LPTAFITWYF  
 10 201 SGYMLGKVLG RTIHVPVPEL LSGGTQDNDL PKEPAKAGTV VAIMLIPMLL  
 251 IFLNTGVSAI ISEKLVSADE TWVQTAKIIG STPIALLISV LVALFVLGRK  
 301 RGESESALEK TVDGAAPVC SVILITGAGG MFGGVLRSAG IGKALADSMA  
 351 DLGIPVLLGC FLVALALRIA QGSATVALTT AAALMAPAVA AAGFTDWQLA  
 401 CIVLATAAGS VGCSHFENDSG FWLVGRLLDM DVPTTLKTWT VNQTLIALIG  
 15 451 FALSALLFAI V\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF140 shows 95.4% identity over a 87aa overlap with an ORF (ORF140a) from strain A of *N.meningitidis*:

20	orfl40.pep	MDGWTQTLAQTLLGISAAAIILILILIVRFRIHALLTVIVSLLTALATGLPTGSIVKD
	orfl40a	MDGWTQTLAQTLLGISAAAIILILILIVKFRHALLTVIVSLLTALATGLPTGSIVND
25		
	orfl40.pep	ILVKNFGGTLGGVALLVGLGAMLERLV
	orfl40a	VLVKNFGGTLGGVALLVGLGAMLRVETSSGGAQSLADALIRMGKRAPFALGVASLIF
30		

The complete length ORF140a nucleotide sequence <SEQ ID 587> is:

1 ATGGACGGCT GGACACAGAC GCTGTCCGCG CAAACCCTGT TGGGCATTTT  
 51 GCGGCGGCA ATCATCCTCA TTCTGATTTT AATCGTCAA TCCGCATCC  
 101 ACGCGCTGCT GACACTGGTC ATCGTCAGCC TGCTGACGGC TTTGGCAACC  
 35 151 GGTTTGCCCA CAGGCAGCAT TGTCACGAC GTACTGGTCA AAAACTTCGG  
 201 CCGCACGCTC GCGGCGGTGG CGCTTCTGGT CCGCCTGGGC GCGATGCTCG  
 251 GACGTTTGGT CGAAACATCC GCGGCGCAC AGTCGCTGGC GGACGCGCTG  
 301 ATCCGGATGT TCGGCGAAA ACGCGCACCG TTCGCGCTGG GCGTTGCCTC  
 351 GCTGATTTTC GGCTTCCCGA TTTCTTTCGA TGCCGACTA ATCGTCATGC  
 40 401 TGCCCATCGT GTTCGCCACC GCACGCGCA TGAACAGGA CGTACTGCC  
 451 TTCGCGCTTG CCTCCATCGG CGCATTTTCC GTCATGCACG TCTTCTGCC  
 501 GCCCCATCCG GCGCCGATTG CCGCTTCCGA ATTTACGGC GCGAACATCG  
 551 GCCAAGTTT GATTTTGGGT CTGCCGACCG CCTTCATCAC ATGGTATTTT  
 601 AGCGGTATA TGCTCGGCAA AGTGTGGGG CGCACCATCC ATGTTCCCGT  
 45 651 TCCCGAATG CTCAGCGGCG GCACGCAAGA CAACGACCTG CCGAAAGAAC  
 701 CTGCCAAAGC AGGAACGGTC GTCGCCATCA TGCTGATTCC CATGCTGCTG  
 751 ATTTCTCTGA ATACCGCGT ATCGGCCCTC ATCAGCGAAA AACTCGTAAG  
 801 TGCGGACGAA ACCTGGGTTC AGACGGCAA AATAATCGGT TCGACACCGA  
 851 TCGCCCTTCT GATTTCCGTA TTGGTCGCAC TGTTGTCTT GGGACGCAA  
 50 901 CGCGGCGAAA GCGGCAGCGC GTTGGAAAA ACCGTGGACG GCGCACTCGC  
 951 CCGCTCTGT TCCGTGATTC TGATTACCGG CGCGGGCGGT ATGTTGGGCG  
 1001 GCGTTTTCG CGCTTCCGGC ATCGGCAAGG CACTCGCCGA CAGCATGGCG  
 1051 GATTTGGGCA TTTCCGTCCT TTTGGGCTGT TTCCTTGTCT CCTTGCACCT  
 1101 GCGTATCGCG CAAGGTTTCG CAACCGTCGC CCGTACCACC GCGCGCGCGC  
 55 1151 TGATGGCTCC TGCCGTTGCC GCGCGCGGT TTACCGACTG GCAGCTCGCC  
 1201 TGTATCGTAT TGGCAACGGC GGCAGGTTTC GTCCGTTGCA GCCACTTCAA  
 1251 CGACTCCGGC TTCTGGCTGG TCGGCCGCGT CTGGACATG GACGTACCGA  
 1301 CCACGCTGAA AACCTGGACG GTCAACCAAA CCTCATCGC ACTCATCGGC  
 1351 TTTGCCTTGT CCGCACTGCT GTTCGCCATC GTCTGA

This encodes a protein having amino acid sequence <SEQ ID 588>:

```

      1 MDGWTQTLISA QTLGISAAA IILILILIVK FRIHALLTLV IVSLLTALAT
    51 GLPTGSIVND VLVKNFGGTL GGVALLVGLG AMLGRLVETS GGAQSLADAL
    101 IRMFGEKRAP FALGVASLIF GFPIFFDAGL IVMLPIVFAT ARRMKQDVLP
    151 FALASIGAFS VMHVFLPPHP GPIAASEFYG ANIGQVLILG LPTAFITWYF
    201 SGYMLGKVLG RTIHVPVPEL LSGGTQDNDL PKEPAKAGTV VAIMLIPMLL
    251 IFLNTGVSAL ISEKLVSAD E TWVQTAKIIG STPIALLISV LVALFVLGRK
    301 RGESESALEK TVDGALAPVC SVILITGAGG MFGGVLRSAG IGKALADSMA
    351 DLGIPVLLGC FLVALALRIA QGSATVALTT AAALMAPAVA AAGFTDWQLA
    401 CIVLATAAGS VGCSEHNDG FWLVGRLLDM DVPTTLKTWT VNQTLIALIG
    451 FALSALLFAI V*
  
```

ORF140a and ORF140-1 show 99.8% identity over a 461aa overlap:

```

    orf140-1.pep MDGWTQTLISAQTLGISAAAIIILILILIVKFRIHALLTLVIVSLLTALATGLPTGSIVND 60
    15 orf140a MDGWTQTLISAQTLGISAAAIIILILILIVKFRIHALLTLVIVSLLTALATGLPTGSIVND 60
    orf140-1.pep IIVKNFGGTLGGVALLVGLGAMLRVETSSGGAQSLADALIRMFGEKRAPFALGVASLIF 120
    20 orf140a VLVKNFGGTLGGVALLVGLGAMLRVETSSGGAQSLADALIRMFGEKRAPFALGVASLIF 120
    orf140-1.pep GFPIFFDAGLIVMLPIVFATARRMKQDVLPFALASIGAFSVMHVFLPPHPGPIAASEFYG 180
    25 orf140a GFPIFFDAGLIVMLPIVFATARRMKQDVLPFALASIGAFSVMHVFLPPHPGPIAASEFYG 810
    orf140-1.pep ANIGQVLILGLPTAFITWYFSGYMLGKVLGRTIHVPPELLSGGTQDNDLPKEPAKAGTV 240
    30 orf140a ANIGQVLILGLPTAFITWYFSGYMLGKVLGRTIHVPPELLSGGTQDNDLPKEPAKAGTV 240
    orf140-1.pep VAIMLIPMLLIFLNTGVSALISEKLVSADETWVQTAKIIGSTPIALLISVLVALFVLGRK 300
    35 orf140a VAIMLIPMLLIFLNTGVSALISEKLVSADETWVQTAKIIGSTPIALLISVLVALFVLGRK 300
    orf140-1.pep RGESESALEKTVDGALAPVCSVILITGAGGMFGGVLRSAGIGKALADSMADLGIPVLLGC 360
    40 orf140a RGESESALEKTVDGALAPVCSVILITGAGGMFGGVLRSAGIGKALADSMADLGIPVLLGC 360
    orf140-1.pep FLVALALRIAQGSATVALTTAAALMAPAVAAAGFTDWQLACIVLATAAGSVGCSEHNDG 420
    45 orf140a FLVALALRIAQGSATVALTTAAALMAPAVAAAGFTDWQLACIVLATAAGSVGCSEHNDG 420
    orf140-1.pep FWLVGRLLDMDVPTTLKTWTVNQTLIALIGFALSALLFAIV 461
    orf140a FWLVGRLLDMDVPTTLKTWTVNQTLIALIGFALSALLFAIV 461
  
```

#### 45 Homology with a predicted ORF from *N.gonorrhoeae*

ORF140 shows 92% identity over a 87aa overlap with a predicted ORF (ORF140ng) from *N.gonorrhoeae*:

```

    orf140.pep MDGWTQTLISAQTLGISAAAIIILILILIVKFRIHALLTLVIVSLLTALATGLPTGSIVKD 60
    50 orf140ng MDGRTQTLISAQTLGISAAAIIILILILIVKFRIALLTLVIASLLTALATGLPTGSIVND 60
    orf140.pep IIVKNFGGTLGGVALLVGLGAMLERLV 87
    orf140ng VLVKNFGGTLGGVALLVGLGAMLRVETSSGGAQSLADALIRMFGEKRAPFAPGVASLIF 120
  
```

55 The complete length ORF140ng nucleotide sequence <SEQ ID 589> was predicted to encode a protein having amino acid sequence <SEQ ID 590>:

```

      1 MDGRTQTLISA QTLGISAAA IILILILIVK FRIRALLTLV IASLLTALAT
    51 GLPTGSIVND VLVKNFGGTL GGVALLVGLG AMLGRLVETS GGAQSLADAL
    101 IRMFGEKRAP FAPGVASLIF GFPIFFDAGL IVMLPIVFAT ARRMKQDVLP
  
```

151 FALASVGAFS VMHVFLPPHP GPIAASEFYG ANIGQVLILG LPTAFITWYF  
 201 SGYMLGKVLG RAIHVPVPEL LSGGTQSDP PKEPAKAGTV VAVMLIPMLL  
 251 IFLNTGVSAL ISEKLVSAD TWVQTAKMIG STPVALLISV LAALLVLGRK  
 301 RGESGSTLEK TVDGALAPAC SVILITGAGG MFGGVLRASG IGKALADSMA  
 5 351 DLGIPVLLGC FLVALALRIA QGSATVALTT AAALMAPAVA AAGFTDWQLA  
 401 CIVLATAAGS VGCSHFND FWLVGRLLSDM DVPTTLKTTWT VNQTLLIAFIG  
 451 FALSALLFAI V\*

Further work revealed a variant gonococcal DNA sequence <SEQ ID 591>:

1 ATGGACGGCC GGACACAGAC GCTGTCCGCG CAAACCTTGT TGGGCATTTC  
 10 51 GCGCGCGGCA ATCATCCTCA TTCTGATTTT AATCGTCAAA TTCCGCATCC  
 101 GCGCGCTGCT GACACTGGTC ATCGCCAGCC TGCTGACGGC TTTGGCAACC  
 151 GGTTCGCCCA CAGGCAGCAT CGTCAACGAC GACTGCTGCA AAAACTTCGG  
 201 CGGCACGCTC GCGGCGTGG CGCTTCTGGT CGGCTGCGG GCAATGCTCG  
 251 GACGTTTGGT AGAAACATCC GCGGCGGCGC AGTCGCTGGC GGACGCGCTG  
 15 301 ATCCGATGT TCGGCGAAAA ACGCGCACCG TTCGCTCCGG GCGTTGCCTC  
 351 GCTGATTTTC GGCTTCCCGA TTTTCTTCGA TGCCGGACTA ATCGTCATGC  
 401 TGCCCATCGT ATTCGCCACC GCACGGCGCA TGAACAGGA CGTACTGCC  
 451 TTCGCGCTTG CCTCCGTCGG CGCATTTTCC GTCATGCACG TCTTCCTGCC  
 501 GCGCCATCCG GCGCCGATTG CCGCTTCCGA ATTTTACGGC GCGAACATCC  
 20 551 GCCAGTTTT GATTTTGGGT CTGCCGACCG CCTTCATCAC ATGGTATTC  
 601 AGCGGCTATA TGCTCGGCAA AGTGTGGGG CGCGCCATCC ATGTTCCCGT  
 651 TCCGAACTG CTCAGCGGCG GCACGCAAGA CAGCGACCCG CCGAAAGAAC  
 701 CTGCCAAAGC AGGAACGTC GTCCGCGTCA TGCTGATTCC CATGCTGCTG  
 25 751 ATTTTCTGA ATACCGCGT ATCAGCCCTC ATCAGCGAAA AACTCGTAAG  
 801 TGCGGACGAA ACTTGGGTTC AGACGGCAAA AATGATCGGT TCGACACCTG  
 851 TCGCCCTTCT GATTTCCGTA TTGGCCGCAC TGTTGGTCTT GGGACGCAAA  
 901 CGCGGCGAAA GCGGCAGCAC GTTGGAAAA ACCGTGGACG GCGCACTCGC  
 951 CCGCGCTGT TCCGTGATTC TGATTACCGG CGCGGGCGGT ATGTTCCGCG  
 30 1001 GCGTTTTCG CGCTTCCGGC ATCGGCAAGG CACTCGCCGA CAGCATGGCG  
 1051 GATTTGGGCA TTCCCGTCTT TTTGGGCTGC TTCTTGTGCG CCTTGGCACT  
 1101 GCGTATCGCG CAAGGTTTCG CAACCGTCGC CCTGACCACA GCGCGCGCGC  
 1151 TGATGGCTCC TGCCGTTGCC GCGCGCGGCT TTACCGACTG GCAGCTCGCC  
 1201 TGTATCGTAT TGGCAACGGC GGCAGGTTTC GTCGGTTGCA GCCACTTCAA  
 1251 CGACTCCGGC TTCTGGCTGG TCGGCCGCTT CTGGATATG GACGTACCGA  
 35 1301 CCACGCTGAA AACCTGGACG GTCAACCAA CCCTCATCGC ATTCATCGGC  
 1351 TTGCGCTGT CCGCACTGCT GTTGCCATC GTCTGA

This corresponds to the amino acid sequence <SEQ ID 592; ORF140ng-1>:

1 MDGRTQTLSA QTLGLISAAA IILILILIVK FRIRALLTLV IASLLTALAT  
 40 51 GLPTGSIVND VLVKNFGGTL GGVALLVGLG AMLGRLVETS GGAQSLADAL  
 101 IRMFGEKRAP FAPGVASLIF GFPIFFDAGL IVMLPIVFAT ARRMKQDVL  
 151 FALASVGAFS VMHVFLPPHP GPIAASEFYG ANIGQVLILG LPTAFITWYF  
 201 SGYMLGKVLG RAIHVPVPEL LSGGTQSDP PKEPAKAGTV VAVMLIPMLL  
 251 IFLNTGVSAL ISEKLVSAD TWVQTAKMIG STPVALLISV LAALLVLGRK  
 301 RGESGSTLEK TVDGALAPAC SVILITGAGG MFGGVLRASG IGKALADSMA  
 45 351 DLGIPVLLGC FLVALALRIA QGSATVALTT AAALMAPAVA AAGFTDWQLA  
 401 CIVLATAAGS VGCSHFND FWLVGRLLDM DVPTTLKTTWT VNQTLLIAFIG  
 451 FALSALLFAI V\*

ORF140ng-1 and ORF140-1 show 96.3% identity over 461aa overlap:

50 orf140ng-1.pep MDGRTQTLSAQTLGLISAAAIILILILIVKFRIRALLTLVIASLLTALATGLPTGSIVND  
 ||| |||||  
 orf140-1 MDGWTQTLSAQTLGLISAAAIILILILIVKFRIRALLTLVIVSLTALATGLPTGSIVND  
 orf140ng-1.pep VLVKNFGGTLGGVALLVGLGAMLGRLVETSSGGAQSLADALIRMFGEKRAPFAPGVASLIF  
 :|||  
 55 orf140-1 IILVKNFGGTLGGVALLVGLGAMLGRLVETSSGGAQSLADALIRMFGEKRAPFALGVASLIF  
 orf140ng-1.pep GFPIFFDAGLIVMLPIVFATARRMKQDVLPPFALASVGAFSVMHVFLPPHPGPIAASEFYG  
 |||  
 60 orf140-1 GFPIFFDAGLIVMLPIVFATARRMKQDVLPPFALASIGAFSVMHVFLPPHPGPIAASEFYG  
 orf140ng-1.pep ANIGQVLILGLPTAFITWYFSGYMLGKVLGRAIHVPVPELLSGGTQSDPPKEPAKAGTV  
 |||  
 orf140-1 ANIGQVLILGLPTAFITWYFSGYMLGKVLGRTHVPVPELLSGGTQDNDLPKEPAKAGTV



orf140ng-1.pep VAVMLIPMLLIFLNTGVSALISEKLVSADETWVQTAKMIGSTPVALLISVLAALLVLGRK  
 orf140-1 VAIMLIPMLLIFLNTGVSALISEKLVSADETWVQTAKIIGSTPIALLISVLVLFVLGRK  
 5 orf140ng-1.pep RGESEGSTLEKTVDGALAPACSVILITGAGGMFGGVLRASGIGKALADSMADLGIPVLLGC  
 orf140-1 RGESEGSALAKTVDGALAPVCSVILITGAGGMFGGVLRASGIGKALADSMADLGIPVLLGC  
 10 orf140ng-1.pep FLVALALRIAQGSATVALTTAAALMAPAVAAAAGFTDWQLACIVLATAAGSVGCSEHNDSDG  
 orf140-1 FLVALALRIAQGSATVALTTAAALMAPAVAAAAGFTDWQLACIVLATAAGSVGCSEHNDSDG  
 15 orf140ng-1.pep FWLVGRLLDMDVPTTLKTTWTVNQTLIAFIGFALSALLFAIV  
 orf140-1 FWLVGRLLDMDVPTTLKTTWTVNQTIALIGFALSALLFAIV

Furthermore, ORF140ng-1 is homologous to an *E.coli* protein:

gi|882633 (U29579) ORF o454 [Escherichia coli] >gi|1789097 (AE000358) o454;  
 This 454 aa ORF is 34% identical (9 gaps) to 444 residues of an approx. 456 aa  
 protein GNTP BACLI SW: P46832 [Escherichia coli] Length = 454  
 Score = 210 bits (529), Expect = 1e-53  
 Identities = 130/384 (33%), Positives = 194/384 (49%), Gaps = 19/384 (4%)  
 20 Query: 88 ETSGGAQSLADALIRMFGEKRAPFAPGVASLIFGFPIFFDAGLIVMLPIVFATARRMKQD 147  
 E SGGA+SLA+ R G+KR A +A+ G P+FFD G I++ PI++ A+ K  
 25 Sbjct: 80 EHSAGAESLANYFSRKLGDKRTIAALTAAFFLGIPVFFDVGFIIAPIIYGFYAKVAKIS 139  
 Query: 148 VLPFALASVGAFSVMHVFLPPHPGPIAASEFYGANIGQVLILGLPTAFITWYFSGYMLGK 207  
 L F L G +HV +PPHPGP+AA+ A+IG + I+G+ + I GY K  
 30 Sbjct: 140 PLKFGFLPVAGIMLTVHVAVPPHPGPVAAAGLLHADIGWLTIIIGIAIS-IPVGVVGYFAAK 198  
 Query: 208 VLGRAIHVPVPELL-----SGGTQSDPPKEPAKAGTVVAVMLIPMLLIFLNTGV 257  
 ++ + + E+L G T+ SD P A V +++IP+ +I T  
 35 Sbjct: 199 IINKRQYAMSVEVLEQMQ LAPASEEGATKLSDKINPPGVA-LVTSILIVIPAIIMAGT-- 255  
 Query: 258 SALISEKLVSADETWVQTAKMIGSTPXXXXXXXXXXXXXGRKRGESEGSTLEKTVDGALA 317  
 +S L+ + T ++IGS +RG S + AL  
 40 Sbjct: 256 ---VSATLMPPSHPLLGLTQLIGSPMVALMIALVLAFWLLALRRGWSLQHTSDIMGSALP 312  
 Query: 318 PACSVILITGAGGMFGGVLRASGIGKALADSMADLGIPVLLGCFVALALRIAQGSXXXX 377  
 A VIL+TGAGG+FG VL SG+GKALA+ + + +P+L F+++LALR +QGS  
 45 Sbjct: 313 TAAVVILVTGAGGVFGKVLVESGVGKALANMLQ MIDLPLLPAAFIISLALRASQGS--AT 370  
 Query: 378 XXXXXXXXXXXXXXXGFTDWQLACIVLATAAGSVGCSEHNDSDGFWLVGRLLDMDVPTTLK 437  
 G Q + LA G +G SH NDSEGFV+V + L + V LK  
 50 Sbjct: 371 VAILTTGGLLSEAVMGLNPIQCVLVTLAACFGGLGASHINDSGFWIVTKYLGLSVADGLK 430  
 Query: 438 TWTVNQTIAFIGFALSALLFAIV 461  
 TWTV T++ F GF ++ ++A++  
 Sbjct: 431 TWTVLTTLGFTGLITWCVWAVI 454

Based on this analysis, including the identification of the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

## 55 Example 71

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 593>:

1 ..GATTTTCGCA TATCGCCCGT GTATCTTTGG GTGCCGCCG CGTTCAAACA  
 51 TTTGCTGTCG CCGTGGGCTG CCGACTCATA CGATGTCGCA CGCTTTGCAG

-356-

5  
101 GCGTATTTTT TGCCGTTATC GGA CTGACTT CCTGCGGCTT TGCCGGTTTC  
151 AACTTTTTTG GCAGACACCA CGGGCGCAC. GTCGTCTGA TTCTCATCGG  
201 CTGTATCGGG CTGATTCCAG TTGCCATT CTCAACCCC GCTGCCGCCG  
251 CCTTGCCGCG CGCCGGACTG GTGCTGCACG GTTATTCTTT GGCTCGCCGG  
301 CGCGTGATG CCGCCTCTTT TCTGCTCGGT ACGGGCTGGA CGTGATGTC  
351 GTTGGCAGCA GCTTATCCGG CAGCATTTGC CCTGATGCTG CCCTGCCCG  
401 TACTGATGTT TTTCCGTCG ..

This corresponds to the amino acid sequence <SEQ ID 594; ORF141>:

10  
1 ..DFGISPVYLW VAAAFKHLSP PWAADSYDVA RFAGVFFAVI GLTSCGFAGF  
51 NFLGRHHGRX VVLILIGCIG LIPVAHFLNP AAAFAAAGL VLHGYSLARR  
101 RVIAASFLLG TGWTLMSLAA AYPAAFALML PLPVLMMFRP ..

Further work revealed the complete nucleotide sequence <SEQ ID 595>:

15  
1 ATGCTGACCT ATACCCCGCC CGATGCCCGC CGGCCCGCCA AAACCCACGA  
51 AAAGCCGTGG CTGCTGCTGT TGATGGCGTT TGCTGGTTG TGGCCCGCGC  
101 TGTTTTCCCA CGATTGTGG AATCCTGACG AACCTGCCGT CTATACCCGC  
151 GTCGAAGCAC TGGCAGGCAG CCCACCCCGC TTGGTTGCCC ATCTGTTCGG  
201 TCAAACCGAT TTCGGCATA CCCCCGTGTA TCTTTGGGT GCCCGCGCGT  
251 TCAAACATTT GCTGTCGCGG TGGGCTGCGG ACTCATACGA TGCCGACGCG  
301 TTTGCAGGCG TATTTTTCG CGTTATCGGA CTGACTTCCT GCGGCTTTGC  
20  
351 CGGTTTCAAC TTTTGGGCA GACACCACGG GCGCAGCGTC GTCCTGATTC  
401 TCATCGGCTG TATCGGGCTG ATTCCAGTTG CCCATTTCCT CAACCCCGCT  
451 GCCGCCGCTT TTGCCCGCGC CGGACTGGTG CTGACGCGTT ATTCTTTGGC  
501 TCGCCGCGCG GTGATTGCGC CCTCTTTTCT GCTCGGTACG GGCTGGACGC  
25  
551 TGATGTCGTT GGCAGCAGCT TATCCGCGAG CATTGCCCCT GATGCTGCCC  
601 TTGCCCGTAC TGATGTTTTT CCGTCCGTGG CAAAGCAGGC GTTTGATGTT  
651 GACGGCAGTC GCCTCACTTG CCTTTGCCCT GCCGCTTATG ACCGTTTACC  
701 CGCTGCTCTT GGCAAAAACG CAGCCCGCGC TGTTCCGCGA ATGGCTCGAC  
751 TATCAGCTTT TCGGTACGTT CGCGCGCGTG CGGCACGTT AGACGGCATT  
30  
801 CAGTTTGTGT TACTATCTGA AAAACCTGCT TTGGTTTGCA TTGCCCGCGC  
851 TGCCCGTGGC GGTTTGGACG GTTTGCCGCA CGCGCCTGTT TTCGACCGAC  
901 TGGGGGATTT TGGGGCTCGT CTGGATGCTT GCCGTTTGG TGCTGCTTGC  
951 CGTCAATCCG CAGCGTTTTT AGGATAACCT CGTCTGGCTG CTTCCGCCCG  
1001 TTGCCCTGTT CGGCGCGGCG CAACTGGACA GCCTGAGGCG CGGCGCGGCG  
35  
1051 GCGTTTGTCA ACTGGTTCGG CATTATGGCG TTCGACTGT TTGCCGTGTT  
1101 CCTGTGGACG GGCTTTTTTC CCATGAATTA CGGCTGGCCC GCCAAGCTTG  
1151 CCGAACGCGC CGCTATTTC AGCCGTATT ATGTTCTGA TATCGATCC  
1201 ATTCGATGG CGGTTGCCGT ACTGTTACA CCCTTGTTGC TGTGGCGAT  
1251 TACCCGAAA AACATACGCG GCAGGCAGGC GGTTACCAAC TGGGCGCAG  
40  
1301 GCGTTACCCT GACCTGGGCT TTGCTGATGA CGCTGTTCTT GCCGTGGCTG  
1351 GACGCGGCGA AAAGCCACGC GCCGTCGTC CGGAGTATGG AGGCATCGCT  
1401 TTCCCGGAA TTGAAACGGG AGCTTTCAGA CGGCATCGAG TGTATCGGCA  
1451 TAGGCGGCG CGACTGCAC ACGCGGATTT TTTGGACGCA GTACGGCACA  
1501 TTGCGGCACC GCGTCGGCGA TGTACAATGC CGCTACCGCA TCGTCTCCT  
45  
1551 GCCCAAAAT GCGGATGCGC CGCAAGGCTG GCAGACGGTT TGGCAGGGTG  
1601 CGCGTCCGCG CAACAAAGAC AGTAAGTTGC CACTGATACG GAAAATCGGG  
1651 GAAAATATAT AA

This corresponds to the amino acid sequence <SEQ ID 596; ORF141-1>:

50  
1 MLTYTPPDAR PPAKTHEKPW LLLMAFAWL WPGVFSHDLW NPDEPAVYTA  
51 VEALAGSPTP LVAHLFGQTD FGIPPVYLWV AAFAKHLSP WAADSYDAAR  
101 FAGVFFAVIG LTSCGFAGFN FLGRHHGRSV VLILIGCIGL IPVAHFLNPA  
151 AAFAAAGLV LHGYSLARRR VIAASFLLGT GWTLMSLAAA YPAAFALMLP  
201 LPVLMFFRPW QSRRLMLTAV ASLAFALPLM TVYPLLLAKT QPALFAQWLD  
251 YHVFGTFGGV RHVQTAFLSF YYLKNLLWFA LPALPLAVWT VCRTRLFSTD  
301 WGILGVVWML AVLVLAVNP QRFQDNLVWL LPPLALFGAA QLDLRLRGA  
55  
351 AFVNWEGIMA EGLFAVFLWT GFFAMNYGWP AKLAERAYF SPYYVPDIDP  
401 IFMAVAVLEF PLWLWAIKTR NIRGRQAVTN WAAGVTLTWA LLMTLFLPWL  
451 DAAKSHAPVV RSMEASLSPE LKRELSDGIE CIGIGGGDLH TRIVWTQYGT  
501 LPHRVGDVQC RYRIVLLPQN ADAPQGWQTV WQGARPRNKD SKFALIRKIG  
551 ENI\*

60 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF141 shows 95.0% identity over a 140aa overlap with an ORF (ORF141a) from strain A of *N.*

*meningitidis:*

[illegible]

The complete length ORF141a nucleotide sequence <SEQ ID 597> is:

25	1	ATGCTGACCT	ATACCCCGCC	CGATGCCCGC	CCGCCCGCCA	AAACCCACGA
	51	AAAGCCGTGG	CTGTTGCTGT	TGATGGCGTT	TGCCTGGTTG	TGGCCCGGCG
	101	TGTTTTCCCA	CGATTTGTGG	AATCCTGACG	AACCTGCCGT	CTATACCGCC
	151	GTCGAAGCAC	TGGCAGGCAG	CCCCACCCCT	TGGTTTGCCC	ATCTGTTCCG
30	201	TCAAATCGAT	TTCGGCATA	CGCCCGTGTA	TCTTTGGGTT	GCCGCCGCGT
	251	TCAAACATT	GCTGTCGCCG	TGGGCTGCCG	ACCCGTAATG	TGCCGCACG
	301	TTTGCCGGCG	TGTTTTTCGC	CGTTGTCCGA	CTGACTTCCT	GCGGCTTTGC
	351	CGGTTTCAAC	TTTTTGGGCA	GACACCACGG	GCGCAGCGTC	GTCCTGATTC
35	401	TCATCGCGTG	TATCGGGCTG	ATTCCGACCG	TACACTTTCT	CAACCCCGCT
	451	GCCGCGCCCT	TTGCCGCGCG	CGGACTGGTG	CTGCACGGTT	ATTTCTTTGGC
	501	TCGCCGCGCG	TGATTGCGCG	CCTCTTTTCT	GTCGGTAGC	GCTTGGACGC
	551	TGATGTCGTT	GGCAGCAGCT	TATCCGGCGG	CATTTGCCCT	GATGCTGCCC
40	601	CTGCCGCTGC	TGATGTTTTT	CCGTCCGTGG	CAAGCAGGC	AGTTTGATGT
	651	GACGGCAGCT	GCCCTCGCTT	CCTTTGCCCT	GCGCCTATG	CGCTTTTACC
	701	CGCTGCTCTT	GGCAAAAACG	CAGCCCGCGC	TGTTCCGCGA	ATGGCTCGAC
	751	GATCACGTTT	TCGGTACGTT	CGCGCGCGTG	GCGCACATTC	ACAGCGGATT
45	801	CAGTTTGTTT	TACTATCTGA	AAAACCTGCT	TGGGTTTGCA	TGTCCTGCGC
	851	TGCCGCTGGC	GGTTTGGACG	GTTTGCCGCA	CGCGCCTGTT	TTCGACCGAC
	901	TGGGGGATT	TGGGCGTCGT	CTGGATGCTT	GCCGTTTTGG	TGCTGCTTGC
	951	CGTCAATCCG	CAGCGTTTTT	AGGATAACCT	CGCTGCGCTG	CTTCGCGCGC
50	1001	TTGCCCTGTT	CGCGCGCGCG	CAACTGGACA	CGCTGAGACG	CGGCGCGCGC
	1051	GCGTTTGTCA	ACTGGTTCGG	CATTATGGCG	TTCGGAAGCT	TGCGCGTGTT
	1101	CCTGTGGACG	GGCTTTTTTC	CCATGAATTA	CGGCTGGCCC	GCCAAGCTTG
	1151	CCGAACGCGC	GCCCTATTTC	AGCCCGTATT	AGGTTCTCTG	TATCGATTCC
55	1201	ATTCGATGG	CGGTTGCCGT	ACTGTTTACA	CCCTTGTTGG	TGTGGGCGAT
	1251	TACCCGCAAA	AACATACGCG	GCAGGCAGGC	GGTTACCAAC	TGGCGGCGAG
	1301	GCGTTACCTT	GACCTGGGCT	TTGCTGATGA	GCGTGTTCCT	GCCGTGGATC
	1351	GACGCGGCGA	AAAGCCACGC	GCCCGTCGTC	CGGAGTATGG	AGGCATCGCT
55	1401	TTCCCCGGAA	TTAAAACGGG	AGCTTTTACA	CGGCATCGAG	TGTATCGACA
	1451	TAGGCGGCGG	CGACCTACAC	ACGCGGATTG	TTTGGACGCA	GTACGGCACA
	1501	TTGCCGCAAC	CGCTCGGCGA	TGTACAATGC	CTGTACGCGA	TGCTCCGCTT
	1551	GCCCCAAAAC	GCGGATGCGC	CGCAAGGCTG	GCAGACGGTC	TGGCAGGGTG
	1601	CGCGCCCGCG	CAACAAAGAC	AGTAAGTTCG	CACTGATACG	GAAAACCGGG
	1651	GAAATATAT	TAAAAACAAC	AGATTGA		

This encodes a protein having amino acid sequence <SEQ ID 598>:

60 1 MLTYTPPDAR PPAKTHEKPW LLLMAFAWL WPGVFSHDLW NPDEPAVYTA  
51 VEALAGSPTP LVAHLFGQID FGIPPVYLWV AA AFKHL LSP WAADPYDAAR  
101 FAGVFFAVVG LTSCGFAGFN FLGRHHGRSV VLILIGICGL IPTVHFLNPA

151 AAFAAAGLV LHGYSLARRR VIAASFLGWTLMSLAAA YPAAFALMLP  
 201 LPVLMFFRPW QSRRLMLTAV ASLAFALPLM TVYPLLAKT QPALFAQWLD  
 251 DHVFGTFGGV RHIQTAFSLF YLKNLLWFA LPALPLAVWT VCRTRLFSTD  
 301 WGILGVVWML AVLVLAVNPQRFQDNLVWLLPPLALFGAAQLDSLRRGAAAFVNWFGIMA  
 351 AFVNWFGIMA FGLFAVFLWT GFFAMNYGWP AKLAERAAAYFSPYYVPDIDPIPMNAVAVLFTPLWLWAI  
 401 IPMAVAVLFT PLWLWAIWK NIRGRQAVTN WAAGVTLTWA LLMTLFLPWL  
 451 DAAKSHAPVV RSMEASLSPE LKRELSDGIE CIDIGGGDLHTRIVWTQYGT  
 501 LPHRVGDVQC RYRIVRLPQN ADAPQGWQTV WQGARPRNKD SKFALIRKTG  
 551 ENILKTTD\*

10 ORF141a and ORF141-1 show 98.2% identity in 553 aa overlap:

orf141a.pep MLTYTPPDARPPAKTHEKPWLLLLMAFAWLWPGVFSHDLWNPDEPAVYTAVEALAGSPTF  
 orf141-1 MLTYTPPDARPPAKTHEKPWLLLLMAFAWLWPGVFSHDLWNPDEPAVYTAVEALAGSPTF  
 15 orf141a.pep LVAHLFGQIDFGIPPVYLWVAAAFKHLSPWAADPYDAARFAGVFFAVVGLTSCGFAGFN  
 orf141-1 LVAHLFGQIDFGIPPVYLWVAAAFKHLSPWAADSYDAARFAGVFFAVVGLTSCGFAGFN  
 20 orf141a.pep FLGRHHGRSVVLILIGCIGLIPTVHFLNPAAAAFAAAGLVLHGYSLARRRVIAASFLG  
 orf141-1 FLGRHHGRSVVLILIGCIGLIPVAHFLNPAAAAFAAAGLVLHGYSLARRRVIAASFLG  
 orf141a.pep GWTLMSLAAAYPAAAFALMLPLPVLMMFFRPWQSRRLMLTAVASLAFALPLMTVYPLLAKT  
 25 orf141-1 GWTLMSLAAAYPAAAFALMLPLPVLMMFFRPWQSRRLMLTAVASLAFALPLMTVYPLLAKT  
 orf141a.pep QPALFAQWLDHVFVGTGGVVRHIQTAFSLFYLLKNLLWFAFPALPLAVWTVCRTRLFSTD  
 orf141-1 QPALFAQWLDHVFVGTGGVVRHVQTAFSLFYLLKNLLWFAFPALPLAVWTVCRTRLFSTD  
 30 orf141a.pep WGILGVVWMLAVLVLAVNPQRFQDNLVWLLPPLALFGAAQLDSLRRGAAAFVNWFGIMA  
 orf141-1 WGILGVVWMLAVLVLAVNPQRFQDNLVWLLPPLALFGAAQLDSLRRGAAAFVNWFGIMA  
 35 orf141a.pep FGLFAVFLWTGFFAMNYGWP AKLAERAAAYFSPYYVPDIDPIPMNAVAVLFTPLWLWAIWK  
 orf141-1 FGLFAVFLWTGFFAMNYGWP AKLAERAAAYFSPYYVPDIDPIPMNAVAVLFTPLWLWAIWK  
 40 orf141a.pep NIRGRQAVTNWAAGVTLT WALLMTLFLPWLDAKSHAPVVRSMELKRELSDGIE  
 orf141-1 NIRGRQAVTNWAAGVTLT WALLMTLFLPWLDAKSHAPVVRSMELKRELSDGIE  
 orf141a.pep CIDIGGGDLHTRIVWTQYGTLP HRVGDVQCRYRIVRLPQNADAPQGWQTVWQGARPRNKD  
 45 orf141-1 CIGIGGGDLHTRIVWTQYGTLP HRVGDVQCRYRIVRLPQNADAPQGWQTVWQGARPRNKD  
 orf141a.pep SKFALIRKTGENI  
 orf141-1 SKFALIRKIGENI  
 50

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF141 shows 95% identity over a 140aa overlap with a predicted ORF (ORF141ng) from *N.gonorrhoeae*:

orf141.pep DFGISPYYLWVAAAFKHLSPWAADSYDVA 30  
 orf141ng WNPAEPVYTAVEALAGSPTPLVAHLFGQIDFGIPPVYLWVAAAFKHLSPWAADSYDVA 126  
 orf141.pep RFAGVFFAVIGLTSCGFAGFNFLGRHHGRXVVLILIGCIGLIPVAHFLNPAAAAFAAAGL 90  
 60 orf141ng RFAGVFFAVIGLTSCGFAGFNFLGRHHGRSVVLILIGCIGLIPVAHFLNPAAAAFAAAGL 186  
 orf141.pep VLHGYSLARRRVIAASFLGWTLMSLAAAYPAAAFALMLPLPVLMMFFRP 140  
 orf141ng VLHGYSLARRRVIAASFLGWTLMSLAAAYPAAAFALMLPLPVLMMFFRPWQSRRLMLTA 246

An ORF141ng nucleotide sequence <SEQ ID 599> was predicted to encode a protein having amino acid sequence <SEQ ID 600>:

```

1  MPSEAVSARP LCEYLLHLAI RFPLLTLMLT YTPPDARPPA KTHEKPWLLL
5  51  LMAFAWLWFG VFSHDLWNPA EPAVYTAVEA LAGSPTPLVA HLFQGQDFGI
101 PPVYLWVAAA FKHLSPWAA HPYDAARFAG VFFAVIGLTS CGFAGFNFLG
151 RHHGRSVVLI HIGCIGLIPV AHFFNPAAAA FAAAGLVLHG YSLARRRVIA
201 ASFLLGTGWT LMSLAAAYPA AFALMLPLEV LMFFRPWQSR RLMLTAVASL
251 AFALPLMTVY PLLLAKTQPA LFAQWLNHYV FGTFGGVRHI QRAFSLFHYL
10  301 KNLLWFAPFG LPLAVWTVCR TRLFSTDWGI LGIVWMLAVL VLLAFNPQRF
351 QDNLVWLLPP LALFGAAQLD SLRRGAAAFV NWFGIMAFGL FAVFLWTGFF
401 AMNYGWPAKL AERAAVFSFY YVPDIDPIM AVAVLFTPLV LWAITRKNIR
451 GRQAVTNWAA GVTLTWALLM TLFPLWLDAA KSHAPVVRSM EASFSPCLKR
501 ELSDGIECIG IGGGDLHTRI VWTQYGTLPV RVGDVRCRYR IVRLPQNADA
551 PQGWQTVWQG ARPRNKDSKF ALIRKIGENI LKTTD*

```

15 Further work revealed the following gonococcal DNA sequence <SEQ ID 601>:

```

1  ATGCTGACCT ATACCCCGCC CGATGCCCGC CCGCCCGCCA AAACCCACGA
5  51  AAAACCGTGG CTGCTGCTGT TGATGGCGTT TGCCTGGCTG TGGCCCGGCG
101 TGTTTCCCA CGATTTGTGG AATCCTGCCG AACCTGCCGT CTATACCGCC
151 GTCGAAGCAC TGGCAGGCAG CCCCACCCCC TTGGTTGCCG ATCTGTTCGG
20  201 TCAAACCGAT TTCGGCATA CCCCCGTGTA TCTTTGGGTT GCCCGCCGAT
251 TCAAACATTT GCTGTCGCCG TGGGCAGCCG ACCCGTATGA TGCCCGACCG
301 TTTGCAGGCG TATTTTTTGC CGTTATCGGA CTGACTTCTT GCGGCTTTGC
351 CGGTTTCAAC TTTTGGGCA GACACCACGG GCGCAGCGTT GTTTTAATCC
401 ATATCGGCTG TATCGGGCTG ATTCCGGTTG CCCATTTCTT CAATCCcgcc
25  451 gccgcgcgct tTGCCGCCGC CGGACTGGTG CTGcagcgt actcgtgCG
501 ACGCCGGCGC GTGATtgccg cctctTtccT GCTCGGTACG GGTGAGACGT
551 TGATGTCGCT GGGGCGAGCT TATCCGCGCG CGTTTGCCTG GATGCTGCCC
601 CTGCCCGTGC TGATGTTTTT CCGTCCGTGG CAAAGCAGGC GTTTGATGTT
651 GACGGCAGTC GCCTCGCTTG CCTTTGCCCT GCCGCTTATG ACCGTTTACC
30  701 CGTGCTCTt gGCAAAACG CAGCCCGCGC TGTTTGCGCA ATGGCTCAAC
751 TATCACGTTT TCGGTACGt cggcgGCGTG CGGCaCaTTC AGAggGCatT
801 Cagtttgttt cactatctga AAaatctgct ttggttcgca ccgccccggc
851 TGCCGCTGGC GGTTTGGACG GTTTGCCGCA CACGCTGTTT TTCGACCGAC
901 TGGGGGATTT TGGGCATTGT CTGGATGCTT GCCGTTTTGG TGCTGCTCGC
35  951 CTTAATCCG CAGCGTTTTT AAGACAACCT CGTCTGGCTG CTGCCCGCCG
1001 TTGCCCTGTT CGGCGCGGCG CAACTGGACA GCCTGAGGCG CGGCGCGGCG
1051 GCTTTTGTC ACTGTTTCGG CATTATGGCG TTCGGGCTGT TTGCCGTGTT
1101 CCTGTGGACG GGCTTTTTTC CCAATGAATTA CGGCTGGCCC GCCAAGCTTG
1151 CCGAACGCGC CGCTACTTC AGCCCGTATT ACGTTCCCGA CATCGATCCC
40  1201 ATTCCGATGG CGGTTGCCGT ACTGTTTACA CCTTTGTGGC TGTGGGCGAT
1251 TACCCGGAAA AACATACGCG GCAGGCAGGC GGTTACCAAC TGGGCGGCAG
1301 GCGTTACCCT GACCTGGGCT TTGCTGATGA CGCTGTTCTT GCCGTGGCTG
1351 GACGCGGCGA AAAGCCACGC GCCCGTCGTC CGGAGTATGG AGGCATCGTT
45  1401 TTCCCGGAA TTAACACGGG AGCTTTCAGA CGGCATCGAG TGTATCGGCA
1451 TAGGCGGCGG CGACCTGCAC ACGCGGATTG TTTGGACGCA GTACGGCACA
1501 TTGCCGCACC GCGTCGGCGA TGTCCGTTGC CGCTACCGTA TCGTCCGCTT
1551 GCCCAAAC GCAGATGCGC CGCAAGGCTG GCAGACGGTC TGGCAGGGTG
1601 CGCGCCCGCG CAACAAAGAC AGTAAGTTTG CACTGATACG GAAATCGGG
1651 GAAATATAT TAAAAACAAC AGATTGA

```

50 This corresponds to the amino acid sequence <SEQ ID 602; ORF141ng-1>:

```

1  MLTYTPPDAR PPAKTHEKPW LLLLMAFAWL WPGVFSHDLW NPAEPAVYTA
5  51  VEALAGSPTP LVAHLFGQTD FGIPPVYLWV AAFAFKHLLSP WAADPYDAAR
101 FAGVFFAVIG LTSCGFAGFN FLGRHHGRSV VLIHIGCIGL IPVAHFLNPA
151 AAFAAAGLV LHGYSLARRR VIAASFLLGT GTLMSLAAA YPAAFALMLP
55  201 LPVLMFFRPW QSRRLMLTAV ASLAFALPLM TVYPLLAKT QFALFAQWLN
251 YHVFGTGGV RHIQRAFSLF HYLKNLLWFA PPGLPLAVWT VCRTRLFSTD
301 WGILGIVWML AVLVLAEFNP QRFQDNLVWL LPPLALFGAA QLSLRRGAA
351 AFVNWFGIMA FGLFAVFLWT GFFAMNYGWP AKLAERAAFY SPYYVDPIDP
401 IPMAVAVLET PLWLWATRKN NIRGRQAVTN WAAGVTLTWA LLMTLFLPWL
60  451 DAAKSHAPVV RSMEASFSPK LKRELSDGIE CIGIGGGDLH TRIVWTQYGT
501 LPHRVGDVRC RYRIVRLPQN ADAPQGWQTV WQGARPRNKD SKFALIRKIG
551 ENILKTTD*

```

ORF141ng-1 and ORF141-1 show 97.5% identity in 553 aa overlap:

```

5  orf141ng-1.pep MLTYTPPDARPPAKTHEKPWLLLLMAFAWLWPGVFSHDLWNPAPAVYTAVEALAGSPTP
   orf141-1      MLTYTPPDARPPAKTHEKPWLLLLMAFAWLWPGVFSHDLWNPDEPAVYTAVEALAGSPTP
10  orf141ng-1.pep LVAHLFGQTDGIPPVYLWVAAAFKHLLSPWAADPYDAARFAGVFFFAVIGLTSCGFAGFN
   orf141-1      LVAHLFGQTDGIPPVYLWVAAAFKHLLSPWAADS YDAARFAGVFFFAVIGLTSCGFAGFN
15  orf141ng-1.pep FLGRHHGRSVVLIHIGCIGLIPVAHFLNPAAAAFAAAGLVLHGYSLARRRVIAASFLLGT
   orf141-1      FLGRHHGRSVVLIHIGCIGLIPVAHFLNPAAAAFAAAGLVLHGYSLARRRVIAASFLLGT
20  orf141ng-1.pep GWTLMSLAAAYPAAFALMLPLPVLMMFFRPWQSRRMLTAVASLAFALPLMTVYPLLLAKT
   orf141-1      GWTLMSLAAAYPAAFALMLPLPVLMMFFRPWQSRRMLTAVASLAFALPLMTVYPLLLAKT
25  orf141ng-1.pep QPALFAQWLNHYHVFQTFGGVRHIQRAFSLFHYLKNLLWFAPPGLPLAVWTVCRTRLFSTD
   orf141-1      QPALFAQWLDYHVFQTFGGVRHVQTAFSLFYLLKNLLWFALPALPLAVWTVCRTRLFSTD
30  orf141ng-1.pep WGILGIVWMLAVLVLLAFNPQRFQDNLVWLLPPLALFGAAQLDSLRRGAAAFVNWFGIMA
   orf141-1      WGILGVVWMLAVLVLLAVNPQRFQDNLVWLLPPLALFGAAQLDSLRRGAAAFVNWFGIMA
35  orf141ng-1.pep FGLFAVFLWTGFFAMNYGWPAKLAERAAYFSPYYVPDIDPIPMAVAVLFTPLWLWAITRK
   orf141-1      FGLFAVFLWTGFFAMNYGWPAKLAERAAYFSPYYVPDIDPIPMAVAVLFTPLWLWAITRK
40  orf141ng-1.pep NIRGRQAVTNWAAGVTLTWALLMTLFLPWLDAAKSHAPVVRSM EASFSP ELKRELS D GIE
   orf141-1      NIRGRQAVTNWAAGVTLTWALLMTLFLPWLDAAKSHAPVVRSM EASFSP ELKRELS D GIE
45  orf141ng-1.pep CIGIGGGDLHTRIVWTQYGTLP HRVGDVRCRYRIVRLPQNADAPQGQWTVWQGARPRNKD
   orf141-1      CIGIGGGDLHTRIVWTQYGTLP HRVGDVQCRYRIVLLPQNADAPQGQWTVWQGARPRNKD
50  orf141ng-1.pep SKFALIRKIGENILKTTDX
   orf141-1      SKFALIRKIGENIX

```

Based on the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 72

45 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 603>:

```

1  ..CAATCCGCCA AATGGTTATC GGGCCAACT CTAGTCGGCA CAGCAATTGG
51 GATACGCGGG CAGATAAAGC TTGGCGGCAA CCTGCATTAC GATATATTTA
101 CCGGCCGCGC ATTGAAAAAG CCCGAATTTT TCCAATCAAG GAAATGGGCA
151 AGCGGTTTTC AGGTAGGCTA TACGTTTAA

```

50 This corresponds to the amino acid sequence <SEQ ID 604; ORF142>:

```

1  ..QSAKWLSGQT LVGTAIGIRG QIKLGGNLHY DIFTGRALKK PEFFQSRKWA
51 SGFQVG YTF*

```

Further work revealed the complete nucleotide sequence <SEQ ID 605>:

```

55 1 ATGGATAATT CGGGTAGTGA GGCGACAGGA AAATACCAAG GAAATATCAC
   51 TTTCTCTGCC GACAATCCTT TGGGACTGAG TGATATGTTC TATGTAAATT

```

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65  
70  
75  
80  
85  
90  
95  
100  
105  
110  
115  
120  
125  
130  
135  
140  
145  
150  
155  
160  
165  
170  
175  
180  
185  
190  
195  
200  
205  
210  
215  
220  
225  
230  
235  
240  
245  
250  
255  
260  
265  
270  
275  
280  
285  
290  
295  
300  
305  
310  
315  
320  
325  
330  
335  
340  
345  
350  
355  
360  
365  
370  
375  
380  
385  
390  
395  
400  
405  
410  
415  
420  
425  
430  
435  
440  
445  
450  
455  
460  
465  
470  
475  
480  
485  
490  
495  
500  
505  
510  
515  
520  
525  
530  
535  
540  
545  
550  
555  
560  
565  
570  
575  
580  
585  
590  
595  
600  
605  
610  
615  
620  
625  
630  
635  
640  
645  
650  
655  
660  
665  
670  
675  
680  
685  
690  
695  
700  
705  
710  
715  
720  
725  
730  
735  
740  
745  
750  
755  
760  
765  
770  
775  
780  
785  
790  
795  
800  
805  
810  
815  
820  
825  
830  
835  
840  
845  
850  
855  
860  
865  
870  
875  
880  
885  
890  
895  
900  
905  
910  
915  
920  
925  
930  
935  
940  
945  
950  
955  
960  
965  
970  
975  
980  
985  
990  
995  
1000

This corresponds to the amino acid sequence <SEQ ID 606; ORF142-1>:

1 MDNSGSEATG KYQGNITFSA DNPLGLSDMF YVNYGRSIGG TPDEESFDGH  
51 RKEGGSNNYA VHYSAPFGKW TWA FNHNGYR YHQA VSGLSE VYDYNKSYN  
101 TDFGFNRLLY RDAKRKTYLG VKLWMRETKS YIDDAELTVQ RRKTAGWLAE  
151 LSHKEYIGRS TADFKLYKXR GTGMKDALRA PEEAFGEGTS RMKIWTASAD  
201 VNTPFQIGKQ LFAYDTSVHA QWNKTPLTSQ DKLAIGGHHT VRGFDGEMSL  
251 SAERGWYWRN DLSWQFKPGH QLYLGADVGH VSGQSAKWLS GQTLVGTAG  
301 IRGQIKLGGN LHYDIFTGRA LKKPEFFQSR KWASGFQVGY TF\*

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF142 shows 88.1% identity over a 59aa overlap with a predicted ORF (ORF142ng) from *N.gonorrhoeae*:

orf142.pep	QSAKWLSGQTLVGTAGIRGQIKLGGNLHY	30
orf142ng	RGWYWRNDLSWQFKPGHQLYLGADVGHVSGQSAKWLSGQTLVGTAGIRGQIKLGGNLHY	313
orf142.pep	DIFTGRALKKPEFFQSRKWASGFQVGYTF	59
orf142ng	DIFTGRALKKPEYFQTKKWVTGFQVGYSF	342

The complete length ORF142ng nucleotide sequence <SEQ ID 607> is:

40  
45  
50  
55  
60  
65  
70  
75  
80  
85  
90  
95  
100  
105  
110  
115  
120  
125  
130  
135  
140  
145  
150  
155  
160  
165  
170  
175  
180  
185  
190  
195  
200  
205  
210  
215  
220  
225  
230  
235  
240  
245  
250  
255  
260  
265  
270  
275  
280  
285  
290  
295  
300  
305  
310  
315  
320  
325  
330  
335  
340  
345  
350  
355  
360  
365  
370  
375  
380  
385  
390  
395  
400  
405  
410  
415  
420  
425  
430  
435  
440  
445  
450  
455  
460  
465  
470  
475  
480  
485  
490  
495  
500  
505  
510  
515  
520  
525  
530  
535  
540  
545  
550  
555  
560  
565  
570  
575  
580  
585  
590  
595  
600  
605  
610  
615  
620  
625  
630  
635  
640  
645  
650  
655  
660  
665  
670  
675  
680  
685  
690  
695  
700  
705  
710  
715  
720  
725  
730  
735  
740  
745  
750  
755  
760  
765  
770  
775  
780  
785  
790  
795  
800  
805  
810  
815  
820  
825  
830  
835  
840  
845  
850  
855  
860  
865  
870  
875  
880  
885  
890  
895  
900  
905  
910  
915  
920  
925  
930  
935  
940  
945  
950  
955  
960  
965  
970  
975  
980  
985  
990  
995  
1000

1001 CGGGGTTTCA GGTGGGTAT TCGTTTGA

This encodes a protein having amino acid sequence &lt;SEQ ID 608&gt;:

1 MDNSGSEATG KYQGNITFSA DNPFGLSDMF YVNYGRSIGG TPDEENFDGH  
 51 RKEGGSNNYA VHYSAPFGKW TWAFNHNGYR YHQAUSGLSE VYDNGKSYN  
 101 TDFGNRLLY RDAKRKTYLS VKLWTRETKS YIDDAELTVQ RRKTTGWLAE  
 151 LSHKGYIGRS TADFELKYKH GTGMKDALRA PEEAFGEGTS RMKIWTASAD  
 201 VNTPFQIGKQ LFAYDTSVHA QWNKTPLTSQ DKLAIGGHHT VRGFDGEMSL  
 251 PAERGWYWRN DLSWQFKPGH QLYLGADVGH VSGQSAKWL SQTLAGTAIG  
 301 IRGQIKLGGN LHYDIFTGRA LKKPEYFQTK KWTGFGVGY SF\*

10 The underlined sequence (aromatic-Xaa-aromatic amino acid motif) is usually found at the C-terminal end of outer membrane proteins.

ORF142ng and ORF142-1 show 95.6% identity over 342aa overlap:

15 orf142-1.pep MDNSGSEATGKYQGNITFSA DNPFGLSDMFYVNYGRSIGGTPDEESFDGHRKEGGSNNYA  
 orf142ng-1 MDNSGSEATGKYQGNITFSA DNPFGLSDMFYVNYGRSIGGTPDEENFDGHRKEGGSNNYA  
 20 orf142-1.pep VHYSAPFGKWTWAFNHNGYRYHQAUSGLSEVYDNGKSYNTDFGNRLLYRDAKRKTYLG  
 orf142ng-1 VHYSAPFGKWTWAFNHNGYRYHQAUSGLSEVYDNGKSYNTDFGNRLLYRDAKRKTYLS  
 25 orf142-1.pep VKLWMRETKSYIDDAELTVQRRKTAGWLAELSHKEYIGRSTADFKLKYKRGTMKDALRA  
 orf142ng-1 VKLWTRETKSYIDDAELTVQRRKTAGWLAELSHKGYIGRSTADFKLKYKHGTGMKDALRA  
 30 orf142-1.pep PEEAFGEGTSRMKIWTASADVNTPFQIGKQLFAYDTSVHAQWNKTPLTSQDKLAIGGHHT  
 orf142ng-1 PEEAFGEGTSRMKIWTASADVNTPFQIGKQLFAYDTSVHAQWNKTPLTSQDKLAIGGHHT  
 35 orf142-1.pep VRGFDGEMSLAERGWYWRNDLSWQFKPGHQLYLGADVGHVSGQSAKWL SGTLVGTAIG  
 orf142ng-1 VRGFDGEMSLPAERGWYWRNDLSWQFKPGHQLYLGADVGHVSGQSAKWL SGTLAGTAIG  
 orf142-1.pep IRGQIKLGGNLHYDIFTGRALKKPEFFQSRKWASGFQVGYTF  
 orf142ng-1 IRGQIKLGGNLHYDIFTGRALKKPEYFQTKKWTGFGVGYSF

In addition, ORF142ng is homologous to the HecB protein of *E. chrysanthemi*:

gi|1772622 (L39897) HecB [Erwinia chrysanthemi] Length = 558  
 Score = 119 bits (295), Expect = 3e-26  
 Identities = 88/346 (25%), Positives = 151/346 (43%), Gaps = 22/346 (6%)  
 40 Query: 2 DMSGSEATGKYQGNITFSA DNPFGLSDMFYVNYGRSIGGTPDEENFDGHRKEGGSNNYAV 61  
 DMSG ++TG+ Q N + + DN FGL+D ++++ G S + + D + G  
 Sbjct: 230 DMSGQKTGEEQLNGSLALDNVFLADQWFI SAGHS---SRFATSHDAESLQAG----- 280  
 45 Query: 62 HYSAPFGKWTWAFNHNGYRYHQAUSGLSEVYDNGKSYNTDFGNRLLYRDAKRKTYLSV 121  
 +S P+G W +N++ RY + G S F +R+++RD KT ++  
 Sbjct: 281 -FSMPYGYWNLGYNYSQSRYNFTINRDFPWHSTGDS DTHRFSLSRVVF RDGTMKTAIAG 339  
 50 Query: 122 KLWTRETKSYIDDAELTVQRRKTGWLAELSHKGYIGRSTADFKLKYKHGTGMKDALRAP 181  
 R +Y++ + L RK + ++H + A F Y G +  
 Sbjct: 340 TFSQRTGNLYLNGSLPSSSRKLSSVSLGVNHSQKLWGGLATFNPTYNRGVRWLGSETDT 399  
 55 Query: 182 EEAFFGEGTSRMKIWTASADVNTPFQIGKQLFAYDTSVHAQWNKTPLTSQDKLAIGGHHTV 241  
 +++ E + WT SA P Y S++ Q++ L ++L +GG ++  
 Sbjct: 400 DKSADPRAEFNKWTLSASYYHPV---TDSITYLGS LYGQYSARALY GSEQTLGGESSI 456  
 60 Query: 242 RGFDGEMSLPAERGWYWRNDLSWQFKP----GHQLYLGA-DVGHVSGQSAKWL SGTLAG 296  
 RGF E RG YWRN+L+WQ G+ ++ A D GH+ + +L G  
 Sbjct: 457 RGF-REQYTSGNRGAYWRNELNWQAWQLPVLGNVTFMAAVDGGHLYNHKQDNSTAASLWG 515



Query: 297 TAIGIRGQIKLGGNLHYDIFTGRALKKPEYFQTKKWVTGFQVGYSF 342  
 A+G+ + L + G + P + Q V G++VG SF  
 Sbjct: 516 GAVGMTVASRW---LSQQVTVGWPIISYPAWLQPDMTMVVGYRVGLSF 558

On the basis of this analysis, it is predicted that the proteins from *N.meningitidis* and  
 5 *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for  
 raising antibodies.

### Example 73

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 609>:

10 1 ATGCGGACGA AATGGTCAGC AGTGAGAAGC TGCTTACTTG GgCGGACACC  
 51 GCCGACATCG ATACCGCTTT GAACCTGTTG TACCGTTTGC AAAAAGCTCGA  
 101 ATTCCTCTAT GGCATGAA ACGGTCATTC AGACGGCATC AATTGwCGG  
 151 ACGAGCAATT GCCGTTGCTG ATGGAACAAT TGTCGGCAG CGGTAAGGCG  
 201 TTATTGGTCG ATCGGAACGG TCTGTATCTT GCCAACGCCA ATTTCCATCA  
 251 TGAGCGGCGG GAAGAGTTGG GGTGTGTTGC GGCAGAAGTC GCACAGATGG  
 15 301 AAAAGAAATA CCGGCTGCTG ATTAAGAAAC AC..

This corresponds to the amino acid sequence <SEQ ID 610; ORF143>:

1 MRTKWSAVRS CTWADTADID TALNLLYRLQ KLEFLYGDEN GHSDGINLXD  
 51 EQLPLLMEQL SGSGKALLVD RNGLYLANAN FHHEAAEELG LLAAEVAQME  
 101 KKYRLLIKNN ..

20 Further work revealed the complete nucleotide sequence <SEQ ID 611>:

1 ATGGAATCAA CACTTTCACT ACAAGCAAAT TTATATCCCC GCCTGACTCC  
 51 TGCCGGTGCA TTTTATGCCG TATCCAGCGA TGCCCCCAGT GCCGGTAAAA  
 101 CTTTGTGCA CAGCCTGTTG AAAGCAGATG CGGACGAAAT GGTGAGCAGT  
 151 GAGAAGCTGC TTACTTGGGC GGACACCGCC GACATCGATA CCGCTTTGAA  
 25 201 CCTGTTGTAC CGTTTGCAA AACTCGAATT CCTCTATGGC GATGAAAACG  
 251 GTCATTCAGA CGGCATCAAT TTGTCGGACG AGCAATTGCC GTTGCTGATG  
 301 GAACAATTGT CCGGCAGCGG TAAGGCGTTA TTGGTCGATC GGAACGGTCT  
 351 GTATCTTGCC AACGCCAATT TCCATCATGA GGCGGCGGAA GAGTTGGGGT  
 401 TGTTGGCGGC AGAAGTCGCA CAGATGGAAA AGAAATACCG GCTGCTGATT  
 451 AAGAACAACC TGTATATCAA CAATAACGCT TGGGGCGTTT GCGATCCTTC  
 501 CGGTCAGAGC GAATTGACAT TTTTCCCAT GTATATCGGT TCAACCAAAT  
 551 TTATTTTGGT TATCGGCGGC ATTCCCGATT TGGGCAAAGA GGCATTTGTT  
 601 ACTTTGGTAA GGATTTTATA CCGCCGTTAC AGCAACCGCG TGTA

This corresponds to the amino acid sequence <SEQ ID 612; ORF143-1>:

35 1 MESTLSLQAN LYPRLTPAGA FYAVSSDAPS AGKTLHLSLL KADADEMVSS  
 51 EKLLTWADTA DIDTALNLLY RLQKLEFLYG DENGHS DGIN LSDEQLPLLM  
 101 EQLSGSGKAL LVDRNGLYLA NANFHHEAAE ELGLLAAEVA QMEKKYRLLI  
 151 KNNLYINNA WGVCPSGQS ELTFFFLYIG STKFILVIGG IPDLGKEAFV  
 201 TLVRILYRRY SNRV\*

40 Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF143 shows 92.4% identity over a 105aa overlap with an ORF (ORF143a) from strain A of *N.meningitidis*:

45 orf143.pep 10 20 30  
 MRTKWSAVRSCTWADTADIDTALNLLYRLQKLEFL  
 :: : ||| ||||| ||||| |||||  
 orf143a GAFYAVSSDXPSAGKTLHLSLLKADADEMVSSEKLLTWAXTADIDTALNLLYRLQKLEFL  
 20 30 40 50 60 70

		40	50	60	70	80	90
5	orf143.pep	YGDENGHS	DGINLXDEQLPL	LMELSGSGK	ALLVDRNGLY	LANANFHHEA	AEELGLLAAE
	orf143a	YGDENGHS	DGINLSDEQLPL	LMELSGSGK	ALLVDRNGLY	LANANFHHEA	AEELGLLAAE
		80	90	100	110	120	130
		100	110				
	orf143.pep	VAQMEKKY	RLLIKNN				
10	orf143a	VAQMEKKY	RLLIKNNLYINN	NAWGVCDPSG	QSELTFEFLY	IGSTKFILVIG	GIPDLGKEA
		140	150	160	170	180	190

The complete length ORF143a nucleotide sequence <SEQ ID 613> is:

	1	ATGGAATCAA	CANTTTC	ACT	ACAAGCAAAT	TTATATC	NCC	GCCTGACTCC
	51	TGCCGGTGCA	TTTTATG	CCG	TATCCAGCGA	TGNCCCC	AGT	GCCGGTAAAA
15	101	CTTGTTTGCA	CAGCCTG	TG	AAAGCGGATG	CGGACGAA	AT	GGTNAGCAGT
	151	GAGAAGCTGC	TTACCTG	GGC	GGANACCGCC	GACATCGA	T	CCGCTTTGAA
	201	CCTGTTGTAC	CGTTTGCA	AAA	AACTCGAATT	CCTCTATG	GC	GATGAAAACG
	251	GTCATTCAGA	CGGCATCA	AT	TTGTCGGACG	AGCAATTG	CC	GTTGCTGATG
	301	GAACAATTGT	CCGGCAGC	GG	TAAGGCGTTA	TTGGTCGA	TC	GGAACGGTCT
20	351	GTATCTTGCC	AACGCCA	AAT	TCCATCATGA	GGCGGCGG	AA	GAGTTGGGGT
	401	TGTTGGCGGC	AGAAGTCG	CA	CAGATGGAAA	AGAAATACC	G	GCTGCNNATT
	451	AAGAACAACC	TGTATATCA	A	CAATAACGCT	TGGGGCGT	TT	GCGATCCTTC
	501	CGGTCAGAGC	GAATTGAC	AT	TTTCCCATT	GTATATCG	GT	TCAACCAAAT
	551	TTATTTTGGT	TATCGGCG	GC	ATCCCGATT	TGGGCAAAGA	G	GCATTGTGTT
25	601	ACTTTGGTAA	GGATNTT	A	CCNCCNGTTA	CAGCAACCG	C	GTGTAAACT
	651	TGGGAGAGAG	GANGGGT	TAT	GCAGCAATTA	TTGA		

This encodes a protein having amino acid sequence <SEQ ID 614>:

	1	MESTXSLQAN	LYXRLTP	PAGAFY	AVSSDXPS	AGKTLHSL	L	KADADEMVSS
	51	EKLLTWAXTA	DIDTALN	LLY	RLQKLEFLY	G	DENGHS	DGINLSDEQLPLLM
30	101	EQLSGSGKAL	LVDNRNGLY	LANANFHHEA	AEELGLLAAE	VAQMEKKY	R	LXIKNNLYINNNAWGVCDPSGQSELTFEFLYIG
	151	KNNLYINNNA	WGVCDPSG	QS	ELTFEFLY	IG	STKFILVIG	GIPDLGKEAFV
	201	TLVRXLXXL	QQPRVKL	GRE	XGLCSNY*			

ORF143a and ORF143-1 show 97.1% identity in 207 aa overlap:

35	orf143a.pep	MESTXSLQANLYXRLTPAGAFYAVSSDXPSAGKTLHSL	LKADADEMVSS	EKLLTWAXTA
	orf143-1	MESTLSLQANLYPRLTPAGAFYAVSSDAPSAGKTLHSL	LKADADEMVSS	EKLLTWADTA
40	orf143a.pep	DIDTALNLLYRLQKLEFLYDENGHS	DGINLSDEQLPLLMELSGSGKALLVDRNGLYLA	
	orf143-1	DIDTALNLLYRLQKLEFLYDENGHS	DGINLSDEQLPLLMELSGSGKALLVDRNGLYLA	
45	orf143a.pep	NANFHHEAAEELGLLAAEVAQMEKKYRLXIKNNLYINN	NAWGVCDPSGQSELTFEFLYIG	
	orf143-1	NANFHHEAAEELGLLAAEVAQMEKKYRLIKNNLYINN	NAWGVCDPSGQSELTFEFLYIG	
	orf143a.pep	STKFILVIGGIPDLGKEAFVTLVRXL		
	orf143-1	STKFILVIGGIPDLGKEAFVTLVRIL		

#### 50 Homology with a predicted ORF from *N.gonorrhoeae*

ORF143 shows 95.5% identity over a 110aa overlap with a predicted ORF (ORF143ng) from *N.gonorrhoeae*:

55	orf143.pep	MRTKWSAVRSCTWADTADIDTALNLLYRLQKLEFLYDENGHS	DGINLXDEQLPLLMEL	60
	orf143ng	MRTKWSAVRSCTWADTADIDTALNLLYRLQKLEFLYDENGHS	DGINLSDEQLPLLMEL	60
	orf143.pep	SGSGKALLVDRNGLYLANANFHHEAAEELGLLAAEVAQMEKKY	RLLIKNN	110
	orf143ng	SGSGKALLVDRNGLYLANANFHHEAAEELGLLAAEVAQMEKKY	RLLIRNNLYINNNAWGV	120

An ORF143ng nucleotide sequence <SEQ ID 615> was predicted to encode a protein having amino acid sequence <SEQ ID 616>:

```

1  MRTKWSAVRS  CSRADTADID  TALNLLYRLQ  KLEFLYGDEN  GHSDGINLSD
51  EQLPLLMEQL  SGSGKALLVD  RNLGLYLANAN  FHHESAEEELG  LLAAEVAQME
101 KKYRLLIRNN  LYINNNAWGV  CDPGQSELT  FFPLYIGSTK  FILVIAGIPD
151 LSKGGICYFG  KDFIPPLQQP  RVKLTGGIM  RQLLISILED  LNNTSTDIIA
201 SAVISTDGLP  MATMLPSHLN  SDRVGAISAT  LLALGSRVQ  ELACGELEQV
251 MIKKGSGYIL  LSQAGKDAVL  VLVAKETGRL  GLILLDAKRA  ARHIAEAI*

```

Further work revealed the following gonococcal DNA sequence <SEQ ID 617>:

```

10      1  ATGGAATCAA  CACTTTCCT  ACAAGCGAAT  TTATATCCCT  GCCTGACTCC
      51  TGCCGGTGCA  TTTTATGCCG  TATCCAGCGA  TGCCCCCAGT  GCCGGTAAAA
     101  CTTTGTTCG  CAGCCTGTTG  AAAGCGGATG  CGGACGAAGT  GGTGAGCAGT
     151  GAGAAGCTGC  TCGCGGCGGA  CACCGCCGAC  ATCGATACCG  CTTTGAACCT
     201  GTTGTAACCG  TTGCAAAAAC  TCGAATTCCT  CTATGGCGAT  GAAAACGGTC
     15      251  ATTCAGACGG  CATCAATTTG  TCGGACGAGC  AATTGCCGTT  GCTGATGGAA
     301  CAATTGTCCG  GCAGCGGTAA  GGCATTATTG  GTCGATCGGA  ACGGTCTGTA
     351  TCTTGCCAAC  GCCAATTTCC  ATCATGAGTC  GGCGGAAGAG  TTGGGGTTGT
     401  TGGCGGCAGA  AGTCGCACAG  ATGGAAAAGA  AATACCGGCT  GCTGATTAGG
     451  AACAACTGT  ATATCAACAA  TAACGCTTGG  GGCGTTTTCG  ATCCTTCCGG
     20      501  TCAGAGCGAA  TTGACATTTT  TCCCATTTGA  TATCGGTTCA  ACCAAATTTA
     551  TTTTGGTTAT  CGCCGCGCATT  CCCGATTGA  GCAAAGAGGC  ATTTGTTACT
     601  TTGGTAAGGA  TTTTATACCG  CCGTTACAGC  AACCGCGTGT  AA

```

This corresponds to the amino acid sequence <SEQ ID 618; ORF143ng-1>:

```

25      1  MESTLSLQAN  LYPCLTPAGA  FYAVSSDAPS  AGKTLRLSLL  KADAEDEVSS
     51  EKLLAADTAD  IDTALNLLYR  LQKLEFLYGD  ENGHSDGINL  SDEQLPLLME
    101  QLSGSGKALL  VDRNGLYLAN  ANFHESAEE  LGLLAAEVAQ  MEKKYRLLIR
    151  NNLYINNNAW  GVCDPSGQSE  LTFFPLYIGS  TKFILVIAGI  PDLSKEAFVT
    201  LVRILYRRYS  NRV*

```

ORF143ng-1 and ORF143-1 show 95.8% identity in 214 aa overlap:

```

30      orf143ng-1.pep  MESTLSLQANLYPCLTPAGAFYAVSSDAPSAGKTLRLSLLKADAEDEVSSSEKLLA-ADTA  59
      orf143-1        MESTLSLQANLYPRLTPAGAFYAVSSDAPSAGKTLRLSLLKADAEDEVSSSEKLLTWADTA  60
      orf143ng-1.pep  DIDTALNLLYRLQKLEFLYGDENGHS DGINLSDEQLPLLMEQLSGSGKALLVDRNGLYLA  119
      orf143-1        DIDTALNLLYRLQKLEFLYGDENGHS DGINLSDEQLPLLMEQLSGSGKALLVDRNGLYLA  120
      orf143ng-1.pep  NANFHESAEEELGLLAAEVAQMEKKYRLLIRNNLYINNNNAWGVCDPSGQSELTFFPLYIG  179
      orf143-1        NANFHESAEEELGLLAAEVAQMEKKYRLLIKNNLYINNNNAWGVCDPSGQSELTFFPLYIG  180
      orf143ng-1.pep  STKFILVIAGIPDLSKEAFVTLVRILYRRYSNRV  213
      orf143-1        STKFILVIGGIPDLGKEAFVTLVRILYRRYSNRV  214

```

45 Based on the presence of the putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 74

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 619>:

```

50      1  ATGACCTTTT  TACAACGTTT  GCAAGGTTTG  GCAGACAATA  AAATCTGTGC
     51  GTTTGCATGG  TTCGTCTGCC  GCCGCTTTGA  TGAAGAACGC  GTACCGCAGr
    101  CGGCGGCAAG  CATGACGTTT  ACGACGCTGC  TGGCACTCGT  CCCCCTGCTG

```

5  
151 ACCGTGATGG TGGCGGTGCG TTCGATTTTC CCCGTGTTCG ACCGCTGGTC  
201 GGATTCGTTC GTCTCCTTCG TCAACCAAAC CATTGTGCCG CA.GGCGCGG  
251 ACATGGTGTT CGACTATATC AATGCGTTCC GCGAGCAGGC GAACCGGCTG  
301 ACGGCAATCG GCAGCGTGAT GCTGGTCGTT ACCTCGCTGA TGCTGATTCC  
351 GACGATAGAC AATACGTTCA ACCGCATCTG G<sub>a</sub>CGGGTCAA wTyCCAGCGT  
401 CCGTGGATG..

This corresponds to the amino acid sequence <SEQ ID 620; ORF144>:

10  
1 MTFLQRLQGL ADNKICAFW FVVRRFDEER VPQXAASMTF TLLALVPVL  
51 TVMVAVASIF PVFDRWSDSF VSEVNQTIVP XGADMVFDYI NAFREQANRL  
101 TAIGSVMLV TSLMLIRTID NTFNRIWRVX XQRPWM...

Further work revealed the complete nucleotide sequence <SEQ ID 621>:

15  
20  
25  
30  
35  
1 ATGACCTTTT TACAACGTTT GCAAGGTTTG GCAGACAATA AAATCTGTGC  
51 GTTTGCATGG TTCGTCGTCC GCCGCTTTGA TGAAGAACGC GTACCGCAGG  
101 CGGCGGCAAG CATGACGTTT ACGACGCTGC TGGCACTCGT CCCCCTGCTG  
151 ACCGTGATGG TGGCGGTGCG TTCGATTTTC CCCGTGTTCG ACCGCTGGTC  
201 GGATTCGTTC GTCTCCTTCG TCAACCAAAC CATTGTGCCG CAGGCGCGCG  
251 ACATGGTGTT CGACTATATC AATGCGTTCC GCGAGCAGGC GAACCGGCTG  
301 ACGGCAATCG GCAGCGTGAT GCTGGTCGTT ACCTCGCTGA TGCTGATTCC  
351 GACGATAGAC AATACGTTCA ACCGCATCTG GCGGGTCAAT TCCCAGCGTC  
401 CGTGGATGAT GCAGTTTCTC GTCTATTGGG CTTTACTGAC GTTCGGGCCG  
451 CTGTCTTTGG GCGTGGGCAT TTCCTTTATG GTCGGCTCGG TACAGGATGC  
501 CGCGCTTGCC TCAGGTGCGC CGCAGTGGTC GGGCGCGTTG CGAACGCGCG  
551 CGACGCTGAC CTTTCATGACG CTTTGTCTGT GGGGGCTGTA CCGCTTCGTG  
601 CCAAACCGCT TCGTTCCCGC GCGGCAGGCG TTTGTCCGGG CTTTGGCAAC  
651 AGCGTTTGTG CTGGAAACCG CGCGCTCCCT CTTCACTTGG TATATGGGCA  
701 ATTTGACGCG CTACCGCTCG ATTTACGGCG CGTTTGCCGC CGTCCGCTTT  
751 TTTCTGTTGT GGCTGAACCT GTTGTGGACG CTGGTCTTGG GCGGCGCGGT  
801 GCTGACTTCT TCACTCTCCT ACTGGCAGG AGAAGCGTTC CGCAGGGGCT  
851 TCGACTCGCG CGGACGGTTT GACGACGTGT TGAAATCCT GCTGCTTCTG  
901 GATGCGGCGC AAAAAGAAGG CAAAGCCTG CCGTTCAGG AGTTCAGACG  
951 GCATATCAAT ATGGGCTACG ACGAGTTGGG CGAGCTTTTG GAAAAGCTGG  
1001 CGCGGCACG CTACATCTAT TCCGGCAGAC AGGGTTGGGT GTTGAAAACG  
1051 GGGGCGGATT CGATTGAGTT GAACGAACTC TTCAAGCTCT TCGTTTACCG  
1101 TCCGTTGCCT TTGGAAAGGG ATCATGTGAA CCAAGCTGTC GATGCGGTAA  
1151 TGACACCGTG TTTGCAGACT TTGAACATGA CGCTGGCAGA GTTTGACGCT  
1201 CAGGCGAAAA AACGGCAGTA G

This corresponds to the amino acid sequence <SEQ ID 622; ORF144-1>:

40  
45  
1 MTFLQRLQGL ADNKICAFW FVVRRFDEER VPQXAASMTF TLLALVPVL  
51 TVMVAVASIF PVFDRWSDSF VSEVNQTIVP QGADMVFDYI NAFREQANRL  
101 TAIGSVMLV TSLMLIRTID NTFNRIWRVN SQRPWMMQFL VYWALTFFG  
151 LSLGVGISFM VGSVQDAALA SGAPQWSGAL RTAATLTFT LLLWGLYRFV  
201 PNRFPARQA FVGALATAFC LETARSLFTW YMGNFDDYRS IYGAFAAVPF  
251 FLLWLNLLWT LVLGGAVLTS SLSYWQGEAF RRGFDSRGRF DDVLKILLLL  
301 DAAQKEGKAL PVQEFRRHIN MGYDELGELL EKLARHGYIY SGRQGWVLKT  
351 GADSIELNEL FKLFPYRPLP VERDHVNQAV DAVMTPLCLQT LNMTLAEFDA  
401 QAKKRQ\*

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF144 shows 96.3% identity over a 136aa overlap with an ORF (ORF144a) from strain A of *N.*

50 *meningitidis*:

55  
orfl44.pep  
orfl44a  
10 20 30 40 50 60  
MTFLQRLQGLADNKICAFW FVVRRFDEERVPQXAASMTFTLLALVPVLTMVAVASIF  
|||||  
MTFLQRLQGLADNKICAFW FVVRRFDEERVPQXAASMTFTLLALVPVLTMVAVASIF  
10 20 30 40 50 60  
70 80 90 100 110 120

orf144.pep PVFDRWSDSFVSFVNQTI VXPQADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID  
 |||||  
 orf144a PVFDRWSDSFVSFVNQTI VXPQADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID  
 |||||  
 70 80 90 100 110 120  
 5  
 130  
 orf144.pep NTFNRIWRVXXQRPWM  
 |||||  
 orf144a NTFNRIWRVNSQRPWMMQFLVYWALLTFGPLSLGVGISFXVGSVQDAALASGAPQWSGAL  
 |||||  
 10 130 140 150 160 170 180

The complete length ORF144a nucleotide sequence <SEQ ID 623> is:

1 ATGACCTTTT TACAACGTTT GCAAGGTTTG GCAGACAATA AAATCTGTGC  
 51 GTTTGCATGG TTCGTCGTCC GCCGCTTTGA TGAAGAACGC GTACCGCAGG  
 101 CGGCGGCAAG CATGACGTTT ACGACACTGC TGGCACTCGT CCCCCTGCTG  
 15 151 ACCGTGATGG TGGCGGTCGC TTCGATTTC CCGGTGTTCC ACCGNTGGTC  
 201 GGATTCGTTT GTCTCCTTCG TCAACCAAAC CATTGTGCCG CAGGGCGCGG  
 251 ACATGGTNTT CGACTATATC AATGCGTTCC GCGAGCAGGC GAACCGGCTG  
 301 ACGGCAATCG GCAGCGTGAT GCTGGTCGTT ACCTCGCNGA TGCTGATTCTG  
 351 GACGATAGAC AATACGTCA ACCGCATCTG GCGGGTCAAT TCCCAGCGTC  
 20 401 CGTGGATGAT GCAGTTTCTC GTCTATTGGG CTTTACTGAC GTTCGGGCGG  
 451 CTGTCTTTGG GCGTGGGCAT TTCTTTTATN GTCGGCTCGG TACAGGATGC  
 501 CGCGCTTGCC TCAGGTGCGC CGCAGTGGTC GGGCGCGTTG CGAACGGCGG  
 551 CGACGCTGAN CTTTCATGACG CTTTGTCTGT GGGGCTGTGA CCGCTNCGTG  
 601 CCAAACCGCT TCGTTCCTCG GCGGCANGCG TTTGTGCGGG CTTTGGAAC  
 25 651 AGCGTTCTGT CTGGAACCG CCGCTTCCCT CTTTACTTGG TATATGGGCA  
 701 ATTTGACGCG CTACCGCTCG ATTTACGNG CGTTGCGCG CGTGCCGTTT  
 751 TTTCTGTTGT GGCTGAACCT GTTGTGGACG CTGGTCTTGG GCGGCGCGGT  
 801 GCTGACTTCT TCACTCTCCT ACTGGCAGG AGAAGCGTTC CGCAGGNGCT  
 851 TCGACTCGCG CGGACGGTTT GACGACGTGT TGAAAATCCT GCTGCTTCTG  
 30 901 GATGCGGCGC AAAAAGAGG CNAAGCCTTG CCTGTCAGG AGTTCAGACG  
 951 GCATATCAAT ATGGGCTACG ACGAGTTGGG CGAGCTTTTG GAAAAGCTGG  
 1001 CGCGGCACGG CTACATCTAT TCCGGCAGAC AGGGTTGGGT GTTGAAACG  
 1051 GGGGCGGATT CGATTGAGTT GAACGAATC TTCAAGCTCT TCGTTTACCG  
 1101 TCCGTTGCCT GTGGAAGGG ATCATGTGAA CCAAGCTGTC GATGCGGTAA  
 35 1151 TGATGCCGTG TTTGCAGACT TTGAACATGA CGCTGGCAGA GTTTGACGCT  
 1201 CAGGCGAAAA AACAGCAGCA ATCTTGA

This encodes a protein having amino acid sequence <SEQ ID 624>:

1 MTFLLQLQGL ADNKICAFW FVRRFDEER VPQAAASMTF TLLALVPVL  
 51 TVMVASIF PVFDRWSDSF VSFVNQTI VXPQADMVFDYI NAFREQANRL  
 40 101 TAIGSVMLV TSXMLIRTID NTFNRIWRVN SQRPWMMQFL VYWALLTFGP  
 151 LSLGVGISFX VGSVQDAALA SGAPQWSGAL RTAATLXFMT LLLWGLYRXV  
 201 PNRFPARXA FVGALATAFC LETARSLFTW YMGNFDDGYRS IYGAFVAVPF  
 251 FLLWLNLLWT LVLGGAVLTS SLSYWQGEAF RRFDSRGRF DDVLKILLLL  
 301 DAAQKEGXAL PVQEFRRHIN MGYDELGELL EKLAHGYIY SGRQGWVLKT  
 45 351 GADSIENEL FKLFFVRPLP VERDHVNQAV DAVMMPCLOT LNMTLAEFDA  
 401 QAKKQQS\*

ORF144a and ORF144-1 show 97.8% identity in 406 aa overlap:

orf144a.pep MTFLLQLQGLADNKICAFW FVRRFDEERVPQAAASMTFTLLALVPVLTVMVASIF  
 |||||  
 50 orf144-1 MTFLLQLQGLADNKICAFW FVRRFDEERVPQAAASMTFTLLALVPVLTVMVASIF  
 orf144a.pep PVFDRWSDSFVSFVNQTI VXPQADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID  
 |||||  
 55 orf144-1 PVFDRWSDSFVSFVNQTI VXPQADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID  
 orf144a.pep NTFNRIWRVNSQRPWMMQFLVYWALLTFGPLSLGVGISFXVGSVQDAALASGAPQWSGAL  
 |||||  
 orf144-1 NTFNRIWRVNSQRPWMMQFLVYWALLTFGPLSLGVGISFXVGSVQDAALASGAPQWSGAL  
 60 orf144a.pep RTAATLXFMTLLLWGLYRXV PNRFPARXAFVVGALATAFCLETARSLFTWYMGNFDDGYRS  
 |||||  
 orf144-1 RTAATLXFMTLLLWGLYRXV PNRFPARXAFVVGALATAFCLETARSLFTWYMGNFDDGYRS

	orf144a.pep	IYGAFAAVFFLLWLNLLWTLVLGGAVLTSSLSYWQGEAFRRXFDSRGRFDDVLKILLLL 	
	orf144-1	IYGAFAAVFFLLWLNLLWTLVLGGAVLTSSLSYWQGEAFRRGFDSRGRFDDVLKILLLL	
5	orf144a.pep	DAAQKEGXALPVQEFRRHINMGYDELGELLEKLARHGYYISGRQGWVLKTGADSIENEL 	
	orf144-1	DAAQKEGKALPVQEFRRHINMGYDELGELLEKLARHGYYISGRQGWVLKTGADSIENEL	
10	orf144a.pep	FKLFVYRPLPVERDHVNQAVDAVMMPCQLTNMTLAEFDAQAKKQQS	408
	orf144-1	FKLFVYRPLPVERDHVNQAVDAVMTPCQLTNMTLAEFDAQAKKRQ	406

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF144 shows 91.2% identity over a 136aa overlap with a predicted ORF (ORF144ng) from

15 *N.gonorrhoeae*:

	orf144.pep	MTFLQRLQGLADNKICAFWFVRRFDEERVPOXAASMTFTLLALVPVLTVMVAVASIF	60
	orf144ng	MTFLQCWQGSADNKICAFWFVIRRFSEERVPOXAASMTFTLLALVPVLTVMVAVASIF	60
20	orf144.pep	PVFDRWSDSFVSFVNQITVPXGADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID	120
	orf144ng	PVFDRWSDSFVSFVNQITVPQGADMVFDYIDAFRDQANRLTAIGSVMLVVTSLMLIRTID	120
25	orf144.pep	NTFNRIWRVXXQRPWM	136
	orf144ng	NAFNRIWRVNTQRPWMMQFLVYWALLTFGPLSLGVGISFMVGSVQDSVLSSGAQQWADAL	180

The complete length ORF144ng nucleotide sequence <SEQ ID 625> is predicted to encode a protein having amino acid sequence <SEQ ID 626>:

30	1	MTFLQCWQGS	ADNKICAFWF	FVIRRFSEER	VPQAAASMTF	TTLLALVPVL
	51	TVMVAVASIF	PVFDRWSDSF	VSEFVNQITVP	QGADMVFDYI	DAFRDQANRL
	101	TAIGSVMLVV	TSLMLIRTID	NAFNRIWRVN	TQRPWMMQFL	VYWALLTFGP
	151	LSLGVGISFM	VGSVQDSVLS	SGAQQWADAL	KTAARLAFMT	LLLWGLYRFV
	201	PNRFVPARQA	FVGALITAF	LETARFLFTW	YMGNFDDGYS	IYGAFAAVFF
35	251	FLWLWLNLLW	LVLGGAVLTS	SLSYWQGEAF	RRGFDSRGRF	DDVLKILLLL
	301	DAAQKEGRTL	SVQEFRRHIN	MGYDELGELL	EKLARYGYIY	SGRQGWVLKT
	351	GADSIELSEL	FKLFVYRPLP	VERDHVNQAV	DAVMTPCQLT	LNMTLAEFDA
	401	QAKKQQQS*				

Further work revealed the following gonococcal DNA sequence <SEQ ID 627>:

40	1	ATGACCTTTT	TACAACGTTG	GCAAGGTTTG	GCGGACAATA	AAATCTGTGC
	51	ATTTCATGCG	TTCGTCATCC	GCCGTTTCAG	TGAAGAGCGC	GTACCGCAGG
	101	CAGCGGCGAG	CATGACGTTT	ACGACACTGC	TGGCACTCGT	CCCCGTACTG
	151	ACCGTAATGG	TCGCGGTCGC	TTCGATTTTC	CCCGTGTTTC	ACCGCTGGTC
	201	GGATTCGTTT	GTCTCCTTCG	TCAACCAAAC	CATTGTGCCG	CAGGGCGCGG
45	251	ATATGGTGTT	CGACTATATC	GACGCATTCC	GCGATCAGGC	AAACCGGCTG
	301	ACCGCCATCG	GCAGCGTGAT	GCTGGTCGTA	ACCTCGCTGA	TGCTGATTCG
	351	GACGATAGAC	AATGCGTTCA	ACCGCATCTG	GCGGGTTAAC	ACGCAACGCC
	401	CCTGGATGAT	GCAGTTCCTC	GTTTATTGGG	CGTTGCTGAC	TTTCGGGCGT
	451	TTGTCTTTGG	GTGTGGGCAT	TTCTTTTATG	GTCTGGGTCG	TTCAAGACTC
50	501	CGTACTCTCC	TCCGGAGCGC	AACAATGGGC	GGACGCGTTG	AAGACGCGCG
	551	CAAGGCTGGC	TTTCATGACG	CTTTTGCTGT	GGGGGCTGTA	CCGCTTCGTG
	601	CCCAACCGCT	TCGTGCCCGC	CCGGCAGGCG	TTTGTCGGAG	CTTTGATTAC
	651	GGCATTCTGC	CTGGAGACGG	CACGTTTCCT	GTTACCTGG	TATATGGGCA
	701	ATTTCCGACG	CTACCGCTCG	ATTTACGGCG	CATTGGCCGC	CGTGCCCTTT
55	751	TTCTGCTGT	GGTTAAACCT	GCTGTGGACG	CTGGTCTTGG	GCGGGGCGGT
	801	GCTGACTTCG	TCGCTGTCTT	ATTGGCAGGG	CGAGGCCTTC	CGCAGGGGAT
	851	TCCGACTCGC	CGGACGTTT	GACGACGTGT	TGAAAATCCT	GCTGCTCTG
	901	GATGCGGCGC	AAAAAGAAGG	CCGAACCTCG	TCCGTTTACG	AGTTCAGACG
	951	GCAATCAAT	ATGGGTACG	ATGAATTGGG	CGAGCTTTTG	GAAAAGCTGG
60	1001	CGCGGTACGG	CTATATCTAT	TCCGGCAGAC	AGGGCTGGGT	TTTGAAAACG
	1051	GGGGCGGATT	CGATTGAGTT	GAGCGAACTC	TTCAAGCTCT	TCGTGTACCG

1101 CCCGTTGCct gtggaAAGGG ATCATGTGAA CCAAGCTGtc gaTGCGGTAA  
 1151 TGAcgccgtG TTTGCAGACT TTGAACATGA CGCTGGCGGA GTTTGACGCT  
 1201 CAGgcgAAAA AACAGCAGCA GTCTTGA

This encodes a variant of ORF144ng, having the amino acid sequence <SEQ ID 628; ORF144ng-

5 1>:

1 MTFLQRWQGL ADNKICAFW FVIRRFSEER VPQAAASMTF TLLALVPVL  
 51 TVMVAVASIF PVFDRWSDSF VSEVNQTIVP QGADMVFDYI DAFRDQANRL  
 101 TAIGSVMLVV TSLMLIRTID NAFNRIWRVN TQRPWMMQFL VYWALLTFGP  
 151 LSLGVGISFM VGSVQDSVLS SGAQQWADAL KTAARLAFMT LLLWGLYRFV  
 201 PNRFPARQA FVGALITAFCL ETARFLFTW YMGNFDGYRS IYGAFAAVPF  
 251 FLLWLNLLWT LVLGGAVLTS SLSYWQGEAF RRGFDSRGRF DDVLKILLLL  
 301 DAAQKEGRTL SVQEFRRHIN MGYDELGELL EKLARYGYIY SGRQGWVLKT  
 351 GADSIELSEL FKL FVYRPLP VERDHVNQAV DAVMTPCLQT LNMTLAEFDA  
 401 QAKKQQQS\*

15 ORF144ng-1 and ORF144-1 show 94.1% identity in 406 aa overlap:

orfl44ng-1.pep MTFLQRWQGLADNKICAFWVIRRFSEERVQAAASMTFTLLALVPVLTVMVAVASIF  
 orfl44-1 MTFLQRLQGLADNKICAFWVVRRFDEERVQAAASMTFTLLALVPVLTVMVAVASIF  
 20 orfl44ng-1.pep PVFDRWSDSFVSEVNQTIVPQGADMVFDYIDAFRDQANRLTAIGSVMLVVTSLMLIRTID  
 orfl44-1 PVFDRWSDSFVSEVNQTIVPQGADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID  
 25 orfl44ng-1.pep NAFNRIWRVNTQRPWMMQFLVYWALLTFGPLSLGVGISFMVGSVQDSVLS SGAQQWADAL  
 orfl44-1 NTFNRIWRVNSQRPWMMQFLVYWALLTFGPLSLGVGISFMVGSVQDAALASGAPQWSGAL  
 30 orfl44ng-1.pep KTAARLAFMTLLLWGLYRFV PNRFPARQAFVGALITAFCL ETARFLFTW YMGNFDGYRS  
 orfl44-1 RTAATLTFTMTLLLWGLYRFV PNRFPARQAFVGALITAFCL ETARSLFTW YMGNFDGYRS  
 35 orfl44ng-1.pep IYGAFAAVPFFLLWLNLLWTLVLGGAVLTSSLSYWQGEAFRRGFDSRGRFDDVLKILLLL  
 orfl44-1 IYGAFAAVPFFLLWLNLLWTLVLGGAVLTSSLSYWQGEAFRRGFDSRGRFDDVLKILLLL  
 40 orfl44ng-1.pep DAAQKEGRTL SVQEFRRHINMGYDELGELLEKLARYGYIYSGRQGWVLKTGADSIELSEL  
 orfl44-1 DAAQKEGKALPVQEFRRHINMGYDELGELLEKLARHGYIYSGRQGWVLKTGADSIELNEL  
 45 orfl44ng-1.pep FKL FVYRPLPVERDHVNQAVDAVMT PCLQTLNMTLAEFDAQAKKQQQS  
 orfl44-1 FKL FVYRPLPVERDHVNQAVDAVMT PCLQTLNMTLAEFDAQAKKRQ

On this basis of this analysis, including the identification of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and  
 45 *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 75

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 629>:

50 1 ..AGACACGCCC GCCGCATCCG CATCGACACC GCCATCAACC CCGAACTGGA  
 51 AGCCCTCGCC GAACACCTCC ACTACCAATG GCAGGGCTTC CTCTGGCTCA  
 101 GCACCGATAT GCGTCAGGAA ATTTCCGCCC TCGTCATCCT GCTGCAACGC  
 151 ACCCGCCGCA AATGGCTGGA TGCCACAGAA CGCCAACACC TGCGCCAAAG  
 201 CCTGCTTGAA ACACGGGAAC ACGGCTGA

This corresponds to the amino acid sequence <SEQ ID 630; ORF146>:

```

1  ..RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTDMRQE ISALVILLQR
51  TRRKWLDAHE RQHLRQSLLE TREHG*

```

Further work revealed the complete nucleotide sequence <SEQ ID 631>:

```

5      1  ATGAACACCT CGCAACGCAA CCGCCTCGTC AGCCGCTGGC TCAACTCCTA
      51  CGAACGCTAC CGCTACCGCC GCCTCATCCA CGCCGTCCGG CTCGGCGGGG
     101  CCGTCCTGTT CGCCACCGCC TCCGCCCGGC TGCTCCACCT CCAACACGGC
     151  GAGTGGATAG GGATGACCGT CTCGTCGTC CTCGGCATGC TCCAGTTTCA
     201  AGGGGCGATT TACTCCAAGG CCGTGGAACG TATGCTCGGC ACGGTCATCG
    10      251  GGCTGGGCGC GGGTTTGGGC GTTTTATGGC TGAACCAGCA TTATTTCAC
     301  GGCAACCTCC TCTTCTACCT CACCGTCGGC ACGGCAAGCG CACTGGCCGG
     351  CTGGGCGGCG GTCGGCAAAA ACGGCTACGT CCCTATGCTG GCAGGGGTGA
     401  CGATGTGTAT GCTCATCGGC GACAACGGCA GCGAATGGCT CGACAGCGGA
     451  CTCATGCGCG CCATGAAAGT CCTCATCGGC GCGGCCATCG CCATCGCCGC
    15      501  CGCCAAACTG CTGCCGCTGA AATCCACACT GATGTGGCGT TTCATGCTTG
     551  CCGACAACCT GGCCGACTGC AGCAAAATGA TTGCCGAAAT CAGCAACGGC
     601  AGGCGCATGA CCGCGAAGC CCTCGAGGAG AACATGGCGA AAATGCGCCA
     651  AATCAACGCA CGCATGGTCA AAAGCCGAG CCATCTCGCC GCCACATCGG
     701  GCGAAAGCCG CATCAGCCCC GCCATGATGG AAGCCATGCA GCACGCCAC
    20      751  CGTAAAATCG TCAACACCAC CGAGCTGCTC CTGACCACCG CCGCCAAGCT
     801  GCAATCTCCC AAATCAACG GCAGCGAAAT CCGGCTGCTT GACCGCCACT
     851  TCACACTGCT CCAAACCGAC CTGCAACAAA CCGTCGCCCT TATCAACGGC
     901  AGACACGCCC GCGCATCCG CATCGACACC GCCATCAACC CCGAACTGGA
     951  AGCCCTCGCC GAACACCTCC ACTACCAATG GCAGGGCTTC CTCTGGCTCA
    25      1001  GCACCAATAT GCGTCAGGAA ATTTCCGCCC TCCTCATCTT GCTGCAACGC
     1051  ACCCGCCGCA AATGGCTGGA TGCCACGAA CGCAACACC TGCGCCAAAG
     1101  CCTGCTTGAA ACACGGGAAC ACGGCTGA

```

This corresponds to the amino acid sequence <SEQ ID 632; ORF146-1>:

```

30      1  MNTSQNRRLV SRWLSYERY RYRRLIHAVR LGGAVLFATA SARLLHLQHG
      51  EWIGMTVFVV LGMLQFQAI YSKAVERMLG TVIGLGAGLG VLWLNQHYFH
     101  GNLLFYLTVG TASALAGWAA VGKNGYVPMI AGLTMCMLIG DNGSEWLDSG
     151  LMRAMNVLIG AATAIAAAKL LPLKSTLMWR FMLADNLADC SKMIAEISNG
     201  RRMTRERLEE NMAKMRQINA RMVKSRSILA ATSGESRIS AMMEAMQHAAH
     251  RKIVNTTELL LTAAKLQSP KLNGSEIRLL DRHFTLLQTD LQQTVALING
    35      301  RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTNMRQE ISALVILLQR
     351  TRRKWLDAHE RQHLRQSLLE TREHG*

```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF146 shows 98.6% identity over a 74aa overlap with an ORF (ORF146a) from strain A of *N.*

40 *meningitidis*:

```

                                     10      20      30
or146.pep                          RHARRIRIDTAINPELEALAEHLHYQWQGF
                                     |||
or146a      KLNGSEIRLLDRHFTLLQTDLQQTVALINGRHARRIRIDTAINPELEALAEHLHYQWQGF
      280      290      300      310      320      330

                                     40      50      60      70
or146.pep      LWLSTDMRQEISALVILLQRTTRRKWLDAHERQHLRQSLLETREHGX
      |||:|||||
or146a      LWLSTNMRQEISALVILLQRTTRRKWLDAHERQHLRQSLLETREHSX
      340      350      360      370

```

The complete length ORF146a nucleotide sequence <SEQ ID 633> is:

```

55      1  ATGAACACCT CGCAACGCAA CCGCCTCGTC AGCCGCTGGC TCAACTCCTA
      51  CGAACGCTAC CGCTACCGCC GCCTCATCCA CGCCGTCCGG CTCGGCGGGG
     101  CCGTCCTGTT CGCCACCGCC TCCGCCCGGC TGCTCCACCT CCAACACGGC
     151  GAGTGGATAG GGATGACCGT CTCGTCGTC CTCGGCATGC TCCAGTTTCA

```



5  
10  
15  
20

```

201 AGGGGCGATT TACTCCAAGG CGGTGGAACG TATGCTCGGC ACGGTCATCG
251 GGCTGGGCGC GGGTTTGGGC GTTTTATGGC TGAACCAGCA TTATTTCCAC
301 GGCAACCTCC TCTTCTACCT CACCGTCGGC ACGGCAAGCG CACTGGCCGG
351 CTGGGCGGCG GTCGGCAAAA ACGGCTACGT CCCTATGCTG GCGGGGCTGA
401 CGATGTGCAT GCTCATCGGC GACAACGGCA GCGAATGGTT CGACAGCGGC
451 CTGATGCGCG CGATGAACGT CCTCATCGGC GCGGCCATCG CCATCGCCGC
501 CGCCAAACTG CTGCCGCTGA AATCCACACT GATGTGGCGT TTCATGCTTG
551 CCGACAACCT GACCGACTGC AGCAAAATGA TTGCCGAAAT CAGCAACGGC
601 AGGCGCATGA CCCGCGAACG CCTCGAAGAG AACATGGCGA AAATGCGCCA
651 AATCAACGCA CGCATGGTCA AAAGCCGCAG CCACCTCGCC GCCACATCGG
701 GCGAAAGCCG CATCAGCCCC GCCATGATGG AAGCCATGCA GCACGCCAC
751 CGTAAAATTG TCAACACCAC CGAGCTGCTC CTGACCACCG CCGCCAAGCT
801 GCAATCTCCC AAACCAACG GCAGCGAAAT CCGGCTGCTT GACCGCCACT
851 TCACACTGCT CCAAACCGAC CTGCAACAAA CCGTCGCCCT TATCAACGGC
901 TGACACGCCC GCCGCATCCG CATCGACACC GCCATCAACC CCGAACTGGA
951 AGCCCTCGCC GAACACCTCC ACTACCAATG GCAGGGCTTC CTCTGGCTCA
1001 GCACCAATAT GCGTCAGGAA ATTTCCGCC TCGTCATCCT GCTGCAACGC
1051 ACCCGCCGCA AATGGCTGGA TGCCCACGAA CGCCAACACC TCGGCCAAAG
1101 CCTGCTTGAA ACACGGGAAC ACAGTTGA

```

20 This encodes a protein having amino acid sequence <SEQ ID 634>:

25

```

1 MNTSQNRNLV SRWLSYERY RYRRLIHAVR LGGAVLFATA SARLLHLQHG
51 EWIGMTVFV LGMLQFQAI YSKAVERMLG TVIGLGAGLG VLWLNQHYFH
101 GNLLFYLTVG TASALAGWAA VGKNGYV PML AGLTMCMLIG DNGSEWFD SG
151 LMRAMNVLIG AAI AIAAAKL LPLKSTLMWR FMLADNLDC SKMIAEISNG
201 RRMTRERLEE NMAKMRQINA RMVKSRLA ATSGESRISP AMMEAMQHAH
251 RKIVNTTELL LTTAAKLQSP KLNSEIRLL DRHFTLLQTD LQQTVALING
301 RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTNMRQE ISALVILLQR
351 TRRKWLDAHE RQHLRQSLLE TREHS*

```

ORF146a and ORF146-1 show 99.5% identity in 374 aa overlap:

30  
35  
40  
45  
50  
55

```

orf146a.pep MNTSQNRNLVSRWLSYERYRYRRLIHAVRLGGAVLFATASARLLHLQHGEWIGMTVFV
|
orf146-1 MNTSQNRNLVSRWLSYERYRYRRLIHAVRLGGAVLFATASARLLHLQHGEWIGMTVFV
|

orf146a.pep LGMLQFQAIYSKAVERMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLTVG TASALAGWAA
|
orf146-1 LGMLQFQAIYSKAVERMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLTVG TASALAGWAA
|

orf146a.pep VGKNGYV PMLAGLTMCLIGDNGSEWFD SGLMRAMNVLIGAAIAIAAAKL LPLKSTLMWR
|
orf146-1 VGKNGYV PMLAGLTMCLIGDNGSEWFD SGLMRAMNVLIGAAIAIAAAKL LPLKSTLMWR
|

orf146a.pep FMLADNLDCSKMIAEISNGRRMTRERLEENMAKMRQINARMVKSRLAATSGESRISP
|
orf146-1 FMLADNLDCSKMIAEISNGRRMTRERLEENMAKMRQINARMVKSRLAATSGESRISP
|

orf146a.pep AMMEAMQHAHRKIVNTTELLLTAAKLQSPKLNSEIRLLDRHFTLLQTDLQQTVALING
|
orf146-1 AMMEAMQHAHRKIVNTTELLLTAAKLQSPKLNSEIRLLDRHFTLLQTDLQQTVALING
|

orf146a.pep RHARRIRIDTAINPELEALAEHLHYQWQGF LWLSTNMRQEISALVILLQRTTRRKWLDAHE
|
orf146-1 RHARRIRIDTAINPELEALAEHLHYQWQGF LWLSTNMRQEISALVILLQRTTRRKWLDAHE
|

orf146a.pep RQHLRQSLLE TREHSX
|
orf146-1 RQHLRQSLLE TREHG
|

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF146 shows 97.3% identity over a 75aa overlap with a predicted ORF (ORF146ng) from  
60 *N.gonorrhoeae*:

```

      orf146.pep                                RHARRIRIDTAINPELEALAEHLHYQWQGF      30
      orf146ng      KLNGSEIRLLDRHFTLLQTDLQQTAALINGRHARRIRIDTAINPELEALAEHLHYQWQGF      364
5      orf146.pep      LWLSTDMRQEISALVILLQRTTRRKWLDAHERQHRLRQSLLTREHG      75
      orf146ng      LWLSTNMRQEISALVIPLQRTTRRKWLDAHERQHRLRQSLLTREHG      409

```

An ORF146ng nucleotide sequence <SEQ ID 635> was predicted to encode a protein having amino acid sequence <SEQ ID 636>:

```

10      1  MSGVRFPSPA  PIPSTDPPSG  SLCFFTFPLQ  TASDMNSSQR  KRLSGRWLNS
      51  YERYRHRLI  HAVRLGGTVL  FATALARLLH  LQHGIEWIGMT  VFVVLGMLQF
      101  QGAIYSNAVE  RMLGTVIGLG  AGLGVLWLNQ  HYFHGNLLFY  LTIGTASALA
      151  GWAAVGKNGY  VPMLAGLTMC  MLIQDNGSEW  LDSGLMRAMN  VLIQAAIAIA
15      201  AAKLLPLKST  LMWRFMLADN  LADCSKMIIE  ISNGRRMTRE  RLEQNMVKMR
      251  QINARMVKSR  SHLAATSGES  RISPSMMEAM  QHAHRKIVNT  TELLLTTAAK
      301  LQSPKLNQSE  IRLLDHFTL  LQTDLQQTAA  LINGRHARRI  RIDTAINPEL
      351  EALAEHLHYQ  WQGFLLSTN  MRQEISALVI  PLQRTTRRWL  DAHERQHRLQ
      401  SLLETREHG*

```

Further work revealed the following gonococcal DNA sequence <SEQ ID 637>:

```

20      1  ATGAACTCCT  CGCAACGCAA  ACGCCTTTCC  GgccGCTGGC  TCAACTCCTA
      51  CGAACGCTac  cGCCaccGCC  GCCTCATACA  TGCCGTGCGG  CTCGGCggaa
      101  ccgtCCTGTT  CGCCACCGCA  CTCGCCCGgc  tACTCCACCT  CCAacacggc
      151  gAATGGATAG  GGatgaCCGT  CTTCGTCGTC  CTCGGCATGC  TCCAGTTCCA
25      201  AGGCgcgatt  tActccaacg  cgttgGAacg  taTgetcgt  acggtcatcg
      251  ggctgGGCGC  GGGTTTGGgc  gTTTATGGC  TGAACCAGCA  TTAttccac
      301  ggcaacCTcc  tttctacct  gaccatcggc  acggcaagcg  cactggccgg
      351  ctGGGCGGCG  GTCGGCAAAA  acggctacgt  ccctatgctg  GCGGGGctgA
      401  CGATGTGCAT  gctcatcggc  gACAACGGCA  GCGAATGGCT  CGACAGCGGC
      451  CTGATGCGCG  CGATGAACGT  CCTCATCGGC  GCCGCCATCG  CCATTGCCGC
30      501  CGCCAAACTG  CTGCCGCTGA  AATCCACACT  GATGTGGCGT  TTCATGCTTG
      551  CCGACAACCT  GGCCGACTGC  AGCAAAATGA  TTGCCGAAAT  CAGCAACGGC
      601  AGGCGTATGA  CGCGCGAACG  TTTGGAGCAG  AATATGGTCA  AAATGCGCCA
      651  AATCAACGCA  CGCATGGTCA  AAAGCCGCAG  CCACCTCGCC  GCCACATCGG
      701  GCGAAAGCCG  CATCAGCCCC  TCCATGATGG  AAGCCATGCA  GCACGCCCAC
35      751  CGCAAATCG  TCAACACCAC  CGAGCTGCTC  CTGACCACCG  CCGCCAAGCT
      801  GCAATCTCCC  AAACCAACG  GCAGCGAAAT  CCGGCTGCTC  GACCGCCACT
      851  TCACACTGCT  CCAAACCGAC  CTGCAACAAA  CCGCCGCCCT  CATCAACGGC
      901  AGACACGCCC  GCCGCATCCG  CATCGACACC  GCCATCAACC  CCGAACTGGA
      951  AGCCCTCGCC  GAACACCTCC  ACTACCAATG  GCAGGGCTTC  CTCTGGCTCA
40      1001  GCACCAATAT  GCGTCAGGAA  ATTCGCGCC  TCGTCATCCT  GCTGCAACGC
      1051  ACCCGCCGCA  AATGGCTGGA  TGCCACGAA  CGCCAACACC  TGCGCCAAG
      1101  CCTGCTTGAA  ACACGGGAAC  ACGGCTGA

```

This corresponds to the amino acid sequence <SEQ ID 638; ORF146ng-1>:

```

45      1  MNSSQRKRLS  GRWLSYERY  RHRRLIHAVR  LGGTVLFATA  LARLLHLQHG
      51  EWIGMTVFVV  LGMLQFQGA  YSNAVERMLG  TVIGLGAGLG  VLWLNQHYFH
      101  GNLLFYLTIG  TASALAGWAA  VGKNGYVPM  AGLTMCMLIG  DNGSEWLD SG
      151  LMRAMNVLI  AAIATAAKL  LPLKSTLMWR  FMLADNLADC  SKMIAEISNG
      201  RRMTRELEQ  NMVKMRQINA  RMVKSRSHLA  ATSGESRISP  SMMEAMQHAH
50      251  RKIVNTTELL  LTAAKLQSP  KLNGSEIRLL  DRHFTLLQTD  LQQTAAALNG
      301  RHARRIRIDT  AINPELEALA  EHLHYQWQGF  LWLSTNMRQE  ISALVILLQR
      351  TRRKWLDAHE  RQHLRQSLE  TREHG*

```

ORF146ng-1 and ORF146-1 show 96.5% identity in 375 aa overlap

```

55      orf146-1.pep      MNTSQRNRLVSRWLSYERYRYRRLIHAVRLGGAVLFATASARLLHLQHGIEWIGMTVFVV
      orf146ng-1      MNSSQRKRLSGRWLSYERYRHRRLIHAVRLGGTVLFATALARLLHLQHGIEWIGMTVFVV
      orf146-1.pep      LGMLQFQGAISKAVRMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLTIGTASALAGWAA
      orf146ng-1      LGMLQFQGAISNAVERMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLTIGTASALAGWAA
60

```

orf146-1.pep VGKNGYVPMLAGLTMCMILGDNSEWLD SGLMRAMNVLIGAAIAIAAAKLLPLKSTLMWR  
 orf146ng-1 VGKNGYVPMLAGLTMCMILGDNSEWLD SGLMRAMNVLIGAAIAIAAAKLLPLKSTLMWR  
 5 orf146-1.pep FMLADNLADCSKMAIEISNGRRMTRERLEENMAKMRQINARMVKSRSRLAATSGESRISP  
 orf146ng-1 FMLADNLADCSKMAIEISNGRRMTRERLEQNMVKMRQINARMVKSRSRLAATSGESRISP  
 10 orf146-1.pep AMMEAMQHAHRKIVNTTELLLTAAKLQSPKLNSEIRLLDRHFTLLQTDLQQTVALING  
 orf146ng-1 SMMEAMQHAHRKIVNTTELLLTAAKLQSPKLNSEIRLLDRHFTLLQTDLQQTAAALING  
 orf146-1.pep RHARRIRIDTAINPELEALAEHLHYQWQGFLWLSTNMQRQISALVILLQRTTRRWLDAHE  
 15 orf146ng-1 RHARRIRIDTAINPELEALAEHLHYQWQGFLWLSTNMQRQISALVILLQRTTRRWLDAHE  
 orf146-1.pep RQHLRQSLLETREHG  
 orf146ng-1 RQHLRQSLLETREHG

20 Furthermore, ORF146ng-1 shows homology with a hypothetical *E.coli* protein:

sp|P33011|YEEA ECOLI HYPOTHETICAL 40.0 KD PROTEIN IN COBU-SBMC INTERGENIC REGION  
 >gi|1736674|gnl|PID|d1016553 (D90838) ORF\_ID:o348#20; similar to [SwissProt  
 Accession Number P33011] [Escherichia coli] >gi|1736682|gnl|PID|d1016560 (D90839)  
 25 ORF\_ID:o348#20; similar to [SwissProt Accession Number P33011] [Escherichia coli]  
 >gi|1788318 (AE000292) f352; 100% identical to fragment YEEA ECOLI SW: P33011 but  
 has 203 additional C-terminal residues [Escherichia coli] Length = 352  
 Score = 109 bits (271), Expect = 2e-23  
 Identities = 89/347 (25%), Positives = 150/347 (42%), Gaps = 21/347 (6%)  
 30 Query: 20 YRHRRLIHAVRLGGTVLFATALARLLHLQHGEWIGMTVFVVLGMLQFQGAISNAVERML 79  
 YRH R++H R+ L + RL + W +T+ V++G + F G + A ER+  
 Sbjct: 15 YRHYRIVHGTRVALAFLTLFLIIRLETFIPESTWPLVTMVVIMGPISFWGNVVPRAFERIG 74  
 35 Query: 80 GTVIGLGAGLGVLWLNQHYFHGNLLFYLTIGTASALAGWAAVVGKNGYVPMLAGLTMCMIL 139  
 GTV+G GL L L L + A L GW A+GK Y +L G+T+ +++  
 Sbjct: 75 GTVLGSILGLIALQLE---LISLPLMLVWCAAMFLCGWALGKPKPYQGLLIGVTLAIVV 131  
 40 Query: 140 GDNGSEWLD SGLMRAMNVLIGXXXXXXXXKLLPLKSTLMWRFMLADNLADCSKMAIEISN 199  
 G E +D+ L R+ +V++G + P ++ + WR LA +L + +++ +  
 Sbjct: 132 GSPTGE-IDTALWRSQDVLGSLAMLFTGIWPQRAFIHWRIQLAKSLTEYNRVYQSAFS 190  
 45 Query: 200 GRRMTRERLEQNMVKMRQINARMVKSRSRLAATSGESRISPSMMEAMQHAHRKIVNXXXX 259  
 + R RLE ++ K+ VK R +A S E+RI S+ E +Q +R +V  
 Sbjct: 191 PNLLERPRLESHLQKLL---TDAVKMRGLIAPASKETRIKSIYEGIQTNRNLCMLEL 247  
 50 Query: 260 XXXXXXXXQSPK---LNGSEIRLLDRHFXXXXXXXXXXAALINGRHARRIRIDTAINPEL 316  
 + LN ++R D AL G +N +  
 Sbjct: 248 QINAYWATRPSHFVLLNAQKLR--DTQHMMQQILLSLVHALYEGNPQPVFANTEKLNDV 305  
 Query: 317 EALAEHL--HYQWQ-----GFLWLSTNMQRQISALVILLQRTTRK 354  
 E L + L H+ + G++WL+ ++ L L+ R RK  
 Sbjct: 306 EELRQLLNHHDLKVVEPTPIYGVWLNMTAHQLELLSNLICRALRK 352

On the basis of this analysis, including the identification of several transmembrane domains in the  
 gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and  
 55 their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 76

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 639>

1 ..GCCGAAGACA CGCGCGTTAC CGCACAGCTT TTGAGCGCGT ACGGCATTCA  
 51 GGGCAAACTC GTCAGTGTGC GCGAACACAA CGAACGCGAG ATGGCGGACA

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65  
70

```

101 AGATTGTCGG CTATCTTTCA GACGGCATGG TTGTGGCACA GGTTCGGAT
151 GCGGGTACGC CGGCCGTGTG CGACCCGGGC GCGAAACTCG CCCGCCGCGT
201 GCGTGAGGCC GGGTTTAAAG TCGTTCCCGT CGTGGGCGCA AC.GCGGTGA
251 TGGCGGCTTT GAGCGTGGCC GGTGTGGAAG GATCCGATTT TTATTTCAAC
301 GGTTTGTAC CGCCGAAATC GGGAGAACGC AGGAAACTGT TTGCCAAATG
351 GGTTCGGGCG GCGTTTCCTA TCGTCATGTT TGAACGCGC CACCGCATCG
401 GTGCAGCGCT TGCCGATATG GCGGAAGTGT TCCCCGAACG CCGATTAATG
451 CTGGCGCGCG AAATTACGAA AACGTTTGAA ACGTCTTAA CGGCACGCGT
501 TGGGGAAATT CAGACGGCAT TGTCTGCCGA CGGCGACCAA TCGCGCGGCG
551 AGATGGTGTT GGTGCTTTAT CCGGCGCAGG ATGAAAAACA CGAAGGCTTG
601 TCCGAGTCCG CGCAAAACAT CATGAAATC CTCACAGCCG AGCTGCCGAC
651 CAAACAGGCG GCGGAGCTTG CTGCCAAAT CACGGGCGAG GGAAAGAAAG
701 CTTTGTACGA T..

```

This corresponds to the amino acid sequence <SEQ ID 640; ORF147>:

15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65  
70

```

1 ..AEDTRVTAQL LSAYGIQGKL VSVREHNERQ MADKIVGYLS DGMVVAQVSD
51 AGTPAVCDPG AKLARRVREA GFKVVPVGA XAVMAALSA GVEGSDFYFN
101 GFVPPKSGER RKLFAKWVRA AFPIVMFETP HRIGALADM AELFPERRLM
151 LAREITKTFE TFLSGTVGEI QTALSADGDO SRGEMVLVLY PAQDEKHEGL
201 SESAQNIMKI LTAELPTKQA AELAAKITGE GKALYD..

```

20 Further work revealed the complete nucleotide sequence <SEQ ID 641>:

25  
30  
35  
40  
45  
50  
55  
60  
65  
70  
75  
80  
85

```

1 ATGTTTCAGA AACATTTGCA GAAAGCCTCC GACAGCGTCG TCGGAGGGAC
51 ATTATACGTG GTTGCCACGC CCATCGGCAA TTGGCGGAC ATTACCTGTC
101 GCGCTTTGGC GGTATTGCAA AAGGCGGACA TCATCTGTGC CGAAGACACG
151 CGCGTTACCG CACAGCTTTT GAGCGCGTAC GGCATTCAGG GCAAACCTCGT
201 CAGTGTGCGC GAACACAACG AACGGCAGAT GGCGGACAAG ATTGTCGGCT
251 ATCTTTTCAGA CGGCATGGTT GTGGCACAGG TTTCCGATGC GGGTACGCGC
301 GCCGTGTGCG ACCCGGGGCG GAAACTCGCC CGCCGCGTGC GTGAGGCCGG
351 GTTTAAAGTC GTTCCCGTCG TGGGCGCAAG CGCGGTGATG GCGGCTTTGA
401 GCGTGGCCGG TGTGGAAGGA TCCGATTTT ATTCAACGG TTTGTACCG
451 CCGAAATCGG GAGAACGCG GAAACTGTT GCCAAATGGG TCGGGGCGGC
501 GTTTCCTATC GTCATGTTG AAACGCCGCA CGCATCGGT GCGACGCTTG
551 CCGATATGGC GGAAGTGTTC CCCGAACGCC GATTAATGCT GCGCGCGGAA
601 ATTACGAAAA CGTTTGAAC GTTCTTAAGC GGCACGGTTG GGGAAATTCA
651 GACGGCATTG TCTGCCGACG GCAACCAATC GCGCGGCGAG ATGGTGTTGG
701 TGCTTTATCC GCGCGAGGAT GAAAAACAG AAGGCTTGTC CGAGTCCGCG
751 CAAAACATCA TGAATCCT CACAGCCGAG CTGCCGACCA AACAGGCGGC
801 GGAGCTTGCT GCCAAATCA CGGGCGAGG AAAGAAAGCT TTGTACGATC
851 TGGCTCTGTC TTGAAAAAC AAATAG

```

This corresponds to the amino acid sequence <SEQ ID 642; ORF147-1>:

40  
45  
50  
55  
60  
65  
70  
75  
80  
85

```

1 MFQKHLQKAS DSVVGGTLYV VATPIGNLAD ITLRALVLQ KADIICAEDT
51 RVTAQLLSAY GIQGLVSVR EHNERQMDK IVGYLSDGMV VAQVSDAGTP
101 AVCDPQAKLA RRVREAGFKV VPVVGASAVM AALSVAGVEG SDFYFNGFVP
151 PKSGERRKLF AKWVRAAFPI VMFETPHRIG ATLADMAELF PERRMLARE
201 ITKTFETFLS GTVGEIQTAL SADGNQSRGE MVLVLYPAQD EKHEGLSESA
251 QNIMKILTAE LPTKQAAELA AKITGEGKKA LYDLALSWKN K*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical protein ORF286 of *E.coli* (accession number U18997)

ORF147 and *E.coli* ORF286 protein show 36% aa identity in 237aa overlap:

50  
55

```

Orf147: 1 AEDTRVTAQLLSAYGIQGLVSVREHNERQMDKIVGYLSDGMVVAQVSDAGTPAVCDPG 60
AEDTR T LL +GI +L ++ +HNE+Q A+ ++ L +G +A VSDAGTP + DPG
Orf286: 43 AEDTRHTGLLLQHFGINARLFAHLDHNEQQKAETLLAKLQEGQNIALVSDAGTPLINDPG 102

Orf147: 61 AKLARRVREXXXXXXXXXXXXXXXXXXXXXXXXXEGSDFYFNGFVPPKSGERRKLFKWVRA 120
L R RE F + GF+P KS RR
Orf286: 103 YHLVRTCREAGIRVVPLPGPCAAITALSAAGLP SDRFCYEGFLPAKSKGRDALKAEAE 162

Orf147: 121 AFPIVMFETPHRIGALADMAELFPERR-LMLAREITKTFETFLSGTVGEIQTALSADGD 179
++ +E+ HR+ +L D+ + E R ++LARE+TKT+ET VGE+ + D +
Orf286: 163 PRTLIFYESTHRLDLSLEDIVAVLGESRYVVLARELTKTWETIHGAPVGEALLAWVKEDEN 222

```

Orf147: 180 QSRGEMVLVLYPAQDEKHEGLSESAQNIMKI LTAELPTKQAAELA AKITGEGKKALY 236  
 + +GEMVL++ + E L A + +L AELP K+AA LAA+I G K ALY  
 Orf286: 223 RRGEMVLIV-EGHKAQEEDLPADALRTLALLQAEPLKKAALAAEIHGVKKNALY 278

### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF147 shows 96.6% identity over a 237aa overlap with ORF75a from strain A of *N. meningitidis*:

					10	20	30
orf147.pep					AEDTRVTAQLLSAYGIQGKLVSVREHNERQ		
orf75a	TLVVATPIGNLADITLRALAVLQKADIICAEDTRVTAQLLSAYGIQGKLVSVREHNERQ						
	20	30	40	50	60	70	
		40	50	60	70	80	90
orf147.pep	MADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPVVGAXAVMAALSVA						
orf75a	MADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPVVGASAVMAALSVA						
	80	90	100	110	120	130	
		100	110	120	130	140	150
orf147.pep	GVEGSDFYFNGFVPPKSGERRKLFKAWVRAAFPIVMFETPHRIGALADMAELFPERRLM						
orf75a	GVAGSDFYFNGFVPPKSGERRKLFKAWVRAAFPIVMFETPHRIGATLADMAELFPERRLM						
	140	150	160	170	180	190	
		160	170	180	190	200	210
orf147.pep	LAREITKTFETFLSGTVGEIQTALSADGDQSRGEMVLVLYPAQDEKHEGLSESAQNIMKI						
orf75a	LAREITKTFETFLSGTVGEIQTALAADGNQSRGEMVLVLYPAQDEKHEGLSESAQNIMKI						
	200	210	220	230	240	250	
		220	230				
orf147.pep	LTAELPTKQAAELA AKITGEGKKALYD						
orf75a	LTAELPTKQAAELA AKITGEGKKALYDLALSWKNKX						
	260	270	280	290			

ORF147a is identical to ORF75a, which includes aa 56-292 of ORF75.

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF147 shows 94.1% identity over a 237aa overlap with a predicted ORF (ORF147ng) from *N.*

*gonorrhoeae*:

orf147.pep					AEDTRVTAQLLSAYGIQGKLVSVREHNERQ	30
orf147ng	TLVVATPIGNLADITLRALAVLQKADIICAEDTRVTAQLLSAYGIQGRVSVREHNERQ					85
orf147.pep	MADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPVVGAXAVMAALSVA					90
orf147ng	MADKIVIGFLSDGLVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPVVGASAVMAALSVA					145
orf147.pep	GVEGSDFYFNGFVPPKSGERRKLFKAWVRAAFPIVMFETPHRIGALADMAELFPERRLM					150
orf147ng	GVAESDFYFNGFVPPKSGERRKLFKAWVRAAFPIVMFETPHRIGATLADMAELFPERRLM					205
orf147.pep	LAREITKTFETFLSGTVGEIQTALSADGDQSRGEMVLVLYPAQDEKHEGLSESAQNIMKI					210
orf147ng	LAREITKTFETFLSGTVGEIQTALAADGNQSRGEMVLVLYPAQDEKHEGLSESAQNAMKI					265
orf147.pep	LTAELPTKQAAELA AKITGEGKKALYD			237		
orf147ng	LAAELPTKQAAELA AKITGEGKKALYDLALSWKNK				300	

An ORF147ng nucleotide sequence <SEQ ID 643> was predicted to encode a protein having amino acid sequence <SEQ ID 644>:

```

1 MSVFQTAFFM FQKHLQKASD SVVGGTLYVV ATPIGNLADI TLRALAVLQK
51 ADIICAEDTR VTAQLLSAYG IQGRLVSVRE HNERQMA DKV IGFLSDGLVV
101 AQVSDAGTPA VCDPGAKLAR RVREAGFKVV PVVGASAVMA ALSVAGVAES
151 DFYFNGFVPP KSGERRKLFA KVVRAAFPVV MFETPHRIGA TLADMAELFP
201 ERLMLAREI TKTFTFSLG TVGEIQTALA ADGNQSRGEM VLVLYPAQDE
251 KHEGLSESAQ NAMKILAAEL PTKQAELAA KITGEGKKAL YDLALSWKNK
301 *

```

10 Further work revealed the following gonococcal DNA sequence <SEQ ID 645>:

```

1 ATGTTTCAGA AACACTTGCA GAAAGCCTCC GACAGCGTCG TCGGAGGGAC
51 ATTATACGTG GTTGCCACGC CCATCGGCAA TTTGGCAGAC ATTACCCTGC
101 GCGCTTTGGC GGTATTGCAA AAGGCGGACA TCATTTGTGC CGAAGACACG
151 CGCGTTACTG CGCAGCTTTT GAGCGCGTAC GGCATTGAGG GCAGGTTGGT
201 CAGTGTGCGC GAACACAACG AGCGGCAGAT GGCGGACAAG GTAATCGGTT
251 TCCTTTTCAGA CGGCCTGGTT GTGGCGCAGG TTTCCGATGC GGGTACGCCG
301 GCCGTGTGCG ACCCGGGCGC GAAACTCGCC CGCCGCGTGC GCGAAGCAGG
351 GTTCAAAGTC GTTCCCCTCG TGGGCGCAAG CGCGGTAATG GCGGCGTTGA
401 GTGTGGCCGG TGTGGCGGAA TCCGATTTT ATTCAACGG TTTGTACCG
451 CCGAAATCGG GCGAACGTAG GAAATTGTTT GCCAAATGGG TCGGGCGGCG
501 ATTTCTGTGC GTCATGTTT AAACGCGCA CCGAATCGGG GCAACGCTTG
551 CCGATATGGC GGAATTGTTT CCCGAACGCC GTCTGATGCT GCGCGCGGAA
601 ATCACGAAAA CGTTTGAAC GTTCTTAAGC GGCACGGTTG GGGAAATTCA
651 GACGGCATTG GCGGCGGACG GCAACCAATC GCGCGGCGAG ATGGTGTGG
701 TGCTTTATCC GCGCAGGAT GAAAAACACG AAGGCTTGTC CGAGTCTGCG
751 CAAATGCGA TGAAATCTT TCGGCGCGAG CTGCCGACCA AGCAGGCGGC
801 GGAGCTTGCC GCAAGATTA CAGGTGAGGG CAAAAGGCT TTGTACGATT
851 TGGCACTGTC GTGAAAAAC AATGA

```

This corresponds to the amino acid sequence <SEQ ID 646; ORF147ng-1>:

```

1 MFQKHLQKAS DSVVGGTLYV VATPIGNLAD ITLRALAVLQ KADIICAEDT
51 RVTAQLLSAY GIQGRLVSVR EHNERQMA DKV VIGFLSDGLV VAQVSDAGTP
101 AVCDPGAKLA RRVREAGFKV VPVVGASAVM AALSVAGVAE SDFYFNGFVP
151 PKSGERRKLF AKVVRAAFPV VMFETPHRIG ATLADMAELF PERRMLARE
201 ITKTFTFSLG TVGEIQTALA ADGNQSRGE MVLVLYPAQD EKHEGLSESA
251 QNAMKILAAE LPTKQAELA AKITGEGKKA LYDLALSWKN K*

```

ORF147ng shows homology to a hypothetical *E. coli* protein:

```

sp|P45528|YRAL_ECOLI HYPOTHETICAL 31.3 KD PROTEIN IN AGAI-MTR INTERGENIC REGION
(F286)
>gi|606086 (U18997) ORF f286 [Escherichia coli]
40 >gi|1789535 (AE000395) hypothetical 31.3 kD protein in agai-mtr intergenic region
[Escherichia coli] Length = 286
Score = 218 bits (550), Expect = 3e-56
Identities = 128/284 (45%), Positives = 171/284 (60%), Gaps = 4/284 (1%)

45 Query: 4 KHLQKASDSVVGGTLYVVATPIGNLADITLRALAVLQKADIICAEDTRVTAQLLSAYGIQ 63
K Q A +S G LY+V TPIGNLADIT RAL VLQ D+I AEDTR T LL +GI
Sbjct: 2 KQHQSADNSQ--GQLYIVPTPIGNLADITQRALEVLQAVDLIAEDTRHTGLLLQHFGIN 59

50 Query: 64 GRLVSVREHNERQMA DKVIGFLSDGLVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPV 123
RL ++ +HNE+Q A+ ++ L +G +A VSDAGTP + DPG L R REAG +VVP+
Sbjct: 60 ARLFALHDHNEQQKAETLLAKLQEGQNIALVSDAGTPLINDPGYHLVRTCREAGIRVVL 119

55 Query: 124 VGASAVMAALSVAGVAESDFYFNGFVPPKSGERRKLFAKVVRAAFPVVFMFETPHRIGATL 183
G A + ALS AG+ F + GF+P KS RR ++ +E+ HR+ +L
Sbjct: 120 PGCAAITALSAAGLPSDRFCYEGFLPAKSKGRRDALKAEAPRTLIFYESTHRLDLSL 179

60 Query: 184 ADMAELFPERR-LMLAREITKTFTFSLGTVGEIQTALAADGNQSRGEMVLVLYPAQDEK 242
D+ + E R ++LARE+TKT+ET VGE+ + D N+ +GEMVL++ +
Sbjct: 180 EDIVAVLGESRYVVLARELTKTWETIHGAPVGELLAWVKEDENRRKGEMVLIV-EGHKAQ 238

Query: 243 HEGLSESAQNAMKILAAELPTKQAELA AKITGEGKKALYDLAL 286

```



5  
10

```

3851 ATCAGuCGCG GCGCGGGTT TTAGCAGCGG CAGCCTTtCA GACGGCATCG
3901 GAGsmAAAwT CCGCGGCCGC GTGctGCATT ACGGCATTCA GGCACGATAC
3951 CCGCGCGgtt tCggCGgAtt CCGCATCGAA CCGCACATCG GCGCAACGcg
4001 ctATTTCGTC CAAAAAGCGG ATTACCGCTA CGAAAACGTC AATATCGCCA
4051 CCCCCGGCCT TGCATTCAAC CGcTACCGCG CGGGCATTAa GGCAGATTAT
4101 TCATTCAAAC CCGCGCAACA CATTTCATC ACGCCTTATT TGAGCCTGTC
4151 CTATACCGAT GCCGCTTCGG GCAAAGTCCG AACACGCGTC AATACCGCGG
4201 TATTGGCTCA GGATTTCGGC AAAACCCGCA GTGCGGAATG GGgCGTAAAC
4251 GCCGAAATCA AAGGTTTCAC GCTGTCCCTC CACGCTGCCG CCGCCAAAGG
4301 CCGCAACTG GAAGCGCAAC ACAGCGCGGG CATCAAATTA GGTACCGCT
4351 GTAA...

```

This corresponds to the amino acid sequence <SEQ ID 648; ORF1>:

15  
20  
25  
30  
35

```

1 MKTTDKRTE THRKAPKTGR IRFXAAYLAI CLSFGILPOA WAGHTYFGIN
51 YQYYRDFAEEN KGFVAVGAKD IEVYNKKGEL VGKSMTKAPM IDFSVVSRRNG
101 VAALVGVQYI VSAVHNGGYN NVDFGAEGXN IXDQXRXTYK IVKRNRYKAG
151 TKGHPYGGDY HMPRLHKXVT DAEPVEMTSY MDGRKYIDQN NYPDRVRIGA
201 GRQYWRSDDED EPNNRESSYH IAS.....GS PMFIYDAQKQ
251 KWLINGVLQT GNPYIGKSNG FQLVRKDFWY DEIFAGDTHS VFYEPRQNGK
301 YSFNDNNGT GKINAKHEHN SLENRLKTRT VQLFNVSLSE TAREPVYHAA
351 GGVNSYRPRL NNGENISFID EGKGELILTS NINQGAGGLY FQGDFTVSPE
401 NNETWQGAGV HISEDSTVTW KVNGVANDRL SKIGKGTLL.....

//

701 .....DKVTAS LTKTDISGNV DLADHAHLNL TGLATLNGNL
751 SANGDTRYTV SHNATQNGNX SLVXNAQATF NQATLNGNTS ASGNASFNLN
801 DHAQVQNGSLT LSGNAKANVS HSALNGNVSL ADKAVFHFES SRFTGQISGG
851 KDTALHLKDS EWTLPSEXEL GNLNLDNATI TLNSAYRHDA AGAQTGSATD
901 APRRRSRRSR RSLIXVTPPT SVESRFNTLT VNGKLNQGT FRFMSELFY
951 RSDKLKLAES SEGTYTLAVN NTGNEPASLE QLTVEGKDN KPLSENLFNT
1001 LQNEHVDAGA W.....

//

1151 .....LDRVFAEDR
1201 RNAVWTSGIR DTKHYRSQDF RAYRQQTDLR QIGMQKNLGS GRVGILFSHN
1251 RTENTFDDGI GNSARLAHGA VEGQYIDRF YIGISAGAGF SSGLSSDGIG
1301 XKXRRRVLHY GIQARYRAGF GFGIEPHIG ATRYFVQKAD YRYENVNIAT
1351 PGLAFNRYRA GIKADYSFKP AQHISITPYL SLSYTDAAAG KVRTRVNTAV
1401 LAQDFGKTRS AEWGVNAEIK GFTLSLHAAA AKGPQLEAQH SAGIKLGYRW
1451 *

```

Further sequencing analysis revealed the complete nucleotide sequence <SEQ ID 649>:

40  
45  
50  
55  
60  
65

```

1 ATGAAACAA CCGACAAACG GACAACCGAA ACACACCGCA AAGCCCCGAA
51 AACCGGCCGC ATCCGCTTCT CGCCTGCTTA CTAGCCATA TGCTGTCTGT
101 TCGGCATTCT TCCCCAAGCC TGGCGGGGAC ACACCTATT TCGGCATCAAC
151 TACCAATACT ATCGCGACTT TGCCGAAAT AAAGGCAAGT TTGCAGTCGG
201 GGCGAAAGAT ATTGAGTTT ACAACAAAA AGGGGAGTTG GTCGGCAAT
251 CAATGACAAA AGCCCCGATG ATTGATTTTT CTGTGGTGTC GCGTAACGGC
301 TGGCGGCAT TGGTGGGCGA TCAATATATT GTGAGCGTGG CACATAACGG
351 CCGCTATAAC AACGTTGATT TTGCTGCGGA AGGAAGAAAT CCCGATCAAC
401 ATCGTTTAC TTATAAAAT GTGAAACGGA ATAATTATA AGCAGGGACT
451 AAAGGCCATC CTTATGGCGG CGATTATCAT ATGCCGCGT TGCAATAATT
501 TGTCACAGAT GCAGAACCTG TTGAAATGAC CAGTTATATG GATGGGCGGA
551 AATATATCGA TCAAAATAAT TACCCTGACC GTGTTCGTAT TGGGGCAGGC
601 AGGCAATATT GCGCATCTGA TGAAGATGAG CCCAATAACC GCGAAAGTTC
651 ATATCATATT GCAAGTGCGT ATTCTTGGCT CGTTGGTGGC AATACCTTTG
701 CACAAAATGG ATCAGGTGGT GGCACAGTCA ACTTAGGTAG TGAATAAATT
751 AAACATAGCC CATATGGTTT TTACCAACA GGAGGCTCAT TTGGCGACAG
801 TGGCTACCA ATGTTTATCT ATGATGCCCA AAAGCAAAAG TGGTTAATTA
851 ATGGGGTATT GCAAACGGGC AACCCCTATA TAGGAAAAAG CAATGGCTTC
901 CAGCTGGTTC GTAAAGATTG GTTCTATGAT GAAATCTTTG CTGGAGATAC
951 CCATTCACTA TTCTACGAAC CACGTCAAAA TGGGAAATAC TCTTTTAAACG
1001 ACGATAATAA TGGCACAGGA AAAATCAATG CCAAACATGA ACACAATTCT
1051 CTGCCATAATA GATTAATAAC ACGAACCGTT CAATTGTTTA ATGTTTCTTT
1101 ATCCGAGACA GCAAGAGAAC CTGTTTATCA TGCTGCAGGT GGTGTCAACA
1151 GTTATCGACC CAGACTGAAT AATGGAGAAA ATATTTCCTT TATTGACGAA
1201 GGAAAAGCGC AATTGATACT TACCAGCAAC ATCAATCAAG GTGCTGGAGG
1251 ATTATATTTT CAAGGAGATT TTACGGTCTC GCCTGAAAT AACGAACTT
1301 GGCAAGGCGC GGGCGTTTAT ATCAGTGAAG ACAGTACCGT TACTTGGAAA
1351 GTAAACGGCG TGGCAAACGA CCGCCTGTCC AAAATCGGCA AAGGCACGCT

```



5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60

```

1401 GCACGTTCAA GCCAAAGGGG AAAACCAAGG CTCGATCAGC GTGGGCGACG
1451 GTACAGTCAT TTTGGATCAG CAGGCAGACG ATAAAGGCAA AAAACAAGCC
1501 TTTAGTGAAA TCGGCTTGGT CAGCGGCAGG GGTACGGTGC AACTGAATGC
1551 CGATAATCAG TTCAACCCCG ACAAACTCTA TTTCGGCTTT CGCGGCGGAC
1601 GTTTGGATTT AAACGGGCAT TCGCTTTCGT TCCACCGTAT TCAAAATACC
1651 GATGAAGGGG CGATGATTGT CAACCACAAT CAAGACAAAG AATCCACCGT
1701 TACCATTACA GGCAATAAAG ATATTGCTAC AACCGGCAAT AACACAGCT
1751 TGGATAGCAA AAAAGAAATT GCCTACAACG GTTGGTTTGG CGAGAAAGAT
1801 ACGACCAAAA CGAACGGGCG GCTCAACCTT GTTTACCAGC CCGCCGAGAG
1851 AGACCGCACC CTGCTGCTTT CCGGCGGAAC AAATTTAAAC GGCAACATCA
1901 CGCAAAACAAA CGGCAAACTG TTTTTCAGCG GCAGACCAAC ACCGCACGCC
1951 TACAATCATT TAAACGACCA TTGGTTCGCA AAAGAGGGCA TTCCTCGCGG
2001 GGAAATCGTG TGGGACAACG ACTGGATCAA CCGCACATTT AAAGCGGAAA
2051 ACTTCCAAAT TAAAGGCGGA CAGGCGGTGG TTTCCCGCAA TGTGCCCCAA
2101 GTGAAAGGCG ATTGGCATTT GAGCAATCAC GCCCAAGCAG TTTTGGTGT
2151 CGCACCGCAT CAAAGCCACA CAATCTGTAC ACGTTCGGAC TGGACGGGTC
2201 TGACAAATTG TGTCGAAAAA ACCATTACCG ACGATAAAGT GATTGCTTCA
2251 TTGACTAAGA CCGACATCAG CGGCAATGTC GATCTTGGCG ATCAGCTCA
2301 TTTAAATCTC ACAGGGCTTG CCACACTCAA CGGCAATCTT AGTGCAAATG
2351 GCGATACACG TTATACAGTC AGCCACAACG CCACCCAAAA CGGCAACCTT
2401 AGCCTCGTGG GCAATGCCCA AGCAACATTT AATCAAGCCA CATTAAACGG
2451 CAACACATCG GCTTCGGGCA ATGCTTCATT TAATCTAAGC GACCACGCCG
2501 TACAAAACGG CAGTCTGACG CTTTCCGGCA ACGCTAAGGC AAACGTAAGC
2551 CATTCGCGAC TCAACGGTAA TGTCTCCCTA GCCGATAAGG CAGTATTCCA
2601 TTTTGAAAGC AGCCGCTTTA CCGGACAAAT CAGCGGCGGC AAGGATACGG
2651 CATTACACTT AAAAGACAGC GAATGGACGC TGCCGTCAGG CACGGAATTA
2701 GGCAATTTAA ACCTTGACAA CGCCACCATT AACTCAATT CCGCTATCG
2751 CCACGATCGC GCAGGGGCGC AAACCGGCAG TCGACAGAT GCGCCGCGCC
2801 GCCGTTCCGC CCGTTCGCGC CGTTCCTTAT TATCCGTAC ACCGCCAAT
2851 TCGGTAGAAT CCGCTTTCAA CACGCTGACG GTAAACGGCA AATTGAACGG
2901 TCAGGGAACA TTCGCTTTA TGTGGAACCT CTTCGGCTAC CGCAGCGACA
2951 AATTGAAGCT GCGGAAAGT TCCGAAGGCA CTACACCTT GCGGTCAC
3001 AATACCGGCA ACGAACCTGC AAGCCTCGAA CAATTGACGG TAGTGGAAG
3051 AAAAGACAAC AAACCGCTGT CCGAAAACCT TAATTTCAAC CTGCAAAACG
3101 AACACGTCGA TCGCGGCGCG TGGCGTTACC AACTCATCCG CAAAGACGGC
3151 GAGTTCGCCC TGCAATATCC GGTCAAAGAA CAAGAGCTTT CCGACAAAT
3201 CGCAAGGCA GAGGCAAAAA AACAGGCGGA AAAAGACAAC GCGCAAGGCC
3251 TTGACGCGCT GATTGCGGCC GGGCGCGATG CCGTCGAAAA GACAGAAAGC
3301 GTTGCCGAAC CGGCCCGGCA GGCAGGCGGG GAAAATGTCT GCATTATGCA
3351 GCGGAGGAA GAGAAAAAAC GGGTGCAGGC GGATAAAGAC ACCGCTTTGG
3401 CGAAACAGCG CGAAGCGGAA ACCCGGCCGG CTACCACCGC CTCCCCCGC
3451 GCCGCGCGCG CCGCGCGGGA TTTGCCGCAA CTGCAACCCC AACCGCAGCC
3501 CCAACCGCAG CGCGACCTGA TCAGCCGTTA TGCCAATAGC GGTTTGAGTG
3551 AATTTTCCGC CAGCCTCAAC AGCGTTTTCG CCGTACAGGA CGAATTAGAC
3601 CGCGTATTTG CCGAAGACCG CCGCAACGCC GTTTGGACAA GCGGCATCCG
3651 GGACACCAAA CACTACCGTT CGCAAGATTT CCGCGCCTAC CGCCAACAAA
3701 CCGACCTGCG CCAATCGGT ATGCAGAAAA ACCTCGGCAG CGGGCGCGTC
3751 GGCATCCTGT TTTGCGACAA CCGGACCGAA AACACCTTCG ACGACGGCAT
3801 CGGCAACTCG GCACGGCTTG CCCACGGCGC CGTTTTCGGG CAATACGGCA
3851 TCGACAGGTT CTACATCGGC ATCAGCGCGG GCGCGGGTTT TAGCAGCGGC
3901 AGCCTTTTCA ACGGCATCGG AGGCAAAATC CGCGCGCGCG TGCTGCATTA
3951 CGGCATTTCG GCACGATACC GCGCCGGTTT CGGCGGATTC GGCATCGAAC
4001 CGCACATCGG CGCAACGCGC TATTTCGTCC AAAAAGCGGA TTACCGCTAC
4051 GAAAACGTCA ATATCGCCAC CCCGGGCTTT GCATTCAACC GCTACCGCGC
4101 GGGCATTAAAG GCAGATTATT CATTCAAACC GCGCAACAC ATTTCCATCA
4151 CGCCTTATTT GAGCCTGTCC TATACCGATG CCGCTTCGGG CAAAGTCCGA
4201 ACACGCGTCA ATACCGCCGT ATTGGCTCAG GATTTGCGCA AAACCGCAG
4251 TCGGGAATGG GCGGTAAACG CCGAAATCAA AGGTTTCAGC CTGTCCCTCC
4301 ACGCTGCCGC CGCCAAAGGC CCGCAACTGG AAGCGCAACA CAGCGCGGGC
4351 ATCAAATTAG GCTACCGCTG GTAA

```

This corresponds to the amino acid sequence <SEQ ID 650; ORF1-1>:

65

```

1 MKTTDKRTE THRKAPKTGR IRFSPAYLAI CLSFGILPQA WAGHTYFGIN
51 YQYRDFAEN KGKFAVGAKE IEVYNKKGEL VGKSMTKAPM IDFSVVSRRG
101 VAALVGDDQYI VSAHNGGYN NVDFGAEGRN PDQHRFTYKI VKRNNYKAGT
151 KGHYPYGGDYH MPRLHKFVTD AEPVEMTSYM DGRKYIDQNN YPDRVRIGAG
201 RQYWRSEDE PNNRESSYHI ASAYSWLVG NTFQNGSGG GTVNLGSEKI
251 KHSYPYGFLPT GSGFGDSGSP MFLYDAQKQK WLINGVLQTG NPYIGKSNFG
301 QLVRKDWFYD EIFAGDTHSV FYEPRQNGKY SFNDNNGTG KINAKHEHNS

```

5	351	LPNRLKTRTV	QLFNVSLSSET	AREPVYHAAG	GVNSYRPLRN	NGENISFIDE
	401	GKGELILTSN	INQGAGGLYF	QGDFTVSPEN	NETWQAGAVH	ISEDSTVTWK
	451	VNGVANDRLS	KIGKGTLHVQ	AKGENQGSIS	VGDGTVILDQ	QADDKGKKQA
	501	FSEIGLVSGR	QTVQLNADNQ	FNPDKLYFGF	RGGRLDLNGH	SLSFHRIQNT
	551	DEGAMIVNHN	GDKESTVTTIT	GNKDIATTGN	NNSLDSKKEI	AYNGWFGGEK
10	601	TTKTNGRLNL	VYQPAEDRT	LLSGGTNLN	GNITQTNGKL	FFSGRPTPHA
	651	YNHLNDHWSQ	KEGIPRGEIV	WDNDWINRTF	KAENFQIKGG	QAVVSRRNAV
	701	VKGDWHLNSH	AQAVFGVAPH	QSHTICTRSD	WTGLDNCVEK	TITDDKVIAS
	751	LTKTDISGNV	DLADHAHLNL	TGLATLNGNL	SANGDTRYTV	SHNATQNGNL
	801	SLVGNAQATF	NQATLNGNTS	ASGNASFNLS	DHAVQNGSLT	LSGNARANVS
15	851	HSALNGNVSL	ADKAVFHFES	SRFTGQISGG	KDATALHLKDS	EWTLPSPGTEL
	901	GNLNLDNATI	TKNASYRHDA	AGAQTGSATD	APRRRSRRSR	RSLLSVTPPT
	951	SVESRFNTLT	VNGKLNQOGT	FRFMSELFYG	RSDKKLKLAES	SEGTYTLAVN
	1001	NTGNEPASLE	QLTVVEGKDN	KPLSENLNFT	LQNEHVDAGA	WRYQLIRKDG
	1051	EFRLHNPVKE	QELSDKLKGA	EAKKQAEKDN	AQSLDALIAA	GRDAVEKTES
20	1101	VAEPARQAGG	ENVGIMQAE	EKKRVQADKD	TALAKQREAE	TRPATTAFFPR
	1151	ARRARRDLPO	LQPQPQPQPQ	RDLISRYANS	GLSEFSAITN	SVFAVQDEL
	1201	RVFAEDRRNA	VWTSGRDTK	HYRSQDFRAY	RQOTDLRQIG	MQKNLGSGRV
	1251	GILFSHNRTE	NFTDDGIGNS	ARLAHGAVFG	QYGIDREFYIG	ISAGAGFSSG
	1301	SLSDGIGGGKI	RRRVLHYGIQ	ARYRAGFGGF	GIEPHIGATR	YFVQKADYRY
	1351	ENVNIATPGL	AFNRYRAGIK	ADYSFKPAQH	ISITPYLSLS	YTDAASCKVR
	1401	TRVNTAVLAQ	DFGKTRSAEW	GVNAEIKGFT	LSLHAAAAGK	PQLEAQHSAG
	1451	IKLGYRW*				

Computer analysis of these sequences gave the following results:

25 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF1 shows 57.8% identity over a 1456aa overlap with an ORF (ORF1a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
30	orf1.pep	MKTTDKRRTTETHRKAPKTGRIRFXAAYLAICLSFGILPQAWAGHTYFGINYYQYRDEAFEN					
	orf1a	MKTTDKRRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQAWAGHTYFGINYYQYRDEAFEN					
		10	20	30	40	50	60
35	orf1.pep	KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVS RNGVAALVGQYIVSVAHNGGYN					
	orf1a	KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVS RNGVAALVGQYIVSVAHNGGYN					
		70	80	90	100	110	120
40	orf1.pep	NVDFGAEGXNIXDQXRXTYKIVKRNNYKAGTKGHPYGGDYHMPRLHKXVTD AEPVEMTSY					
	orf1a	NVDFGAEGXN-PDQHRFSYQIVKRNNYKPDNS-HPYNGDXHMPRLHKFVTD AEPVEMTSD					
		130	140	150	160	170	180
45	orf1.pep	MDGRKYIDQNNYPDRVRIGAGRQYWRSDEDEP-----NN-----					
	orf1a	MRGNTYS DKEYPERVRIGSGHHYWR YDDDKHGDLSYSGAWLIGGNTHMQGWGNGGVXSL					
50		180	190	200	210	220	230
	orf1.pep	----RESSYH-----IA-----SGSPMFIYDAQKQKWLINGVLQTGNPYIGKSNGFQLVRK					
	orf1a	SGDVRHANDYGPMPIAGAAGDSGSPMFIYDKTNKWL LNGVLQTGYPPYSGRENGFQLIRK					
55		240	250	260	270	280	290
	orf1.pep	DW FYDEIFAGDTHSVFYEP RQNGKYSFNDDNNGTGKINAKHEHNSLPNRLKTRTVQLFNV					
	orf1a	DW FYDDIYRGDTHTVXFEP RSNGHFSFTSNNNGTGTVTETNEKVSNP-KLKVQTVRLFDE					
60		300	310	320	330	340	350
		330	340	350	360	370	380

[illegible]

orf1.pep -----  
 5 orf1a LDALIAAGRDAAEKTESVAEPARXAGGENVGIMQAEEEKRVQADKDSALAKQREAE TRP  
 1080 1090 1100 1110 1120 1130  
 760  
 orf1.pep -----LDR  
 10 orf1a XTAFPRARXARRDLFPQPQPQPQPDLDLXSRYANSGLSEFSATLNSVFAVQDELDR  
 1140 1150 1160 1170 1180 1190  
 770 780 790 800 810 820  
 15 orf1.pep VFAEDRRNAVWTS GIRD TKHYRSQDFRAYRQQTDLRQIGMQKNLGS GRV GILF SHNR TEN  
 orf1a VFAEDRRNAVWTSXIRXTKHYRSQDFRAYRQQTDLRQIGMQKNLGS GRV GILF SHNR TEN  
 1200 1210 1220 1230 1240 1250  
 20 orf1.pep 830 840 850 860 870 880  
 TFDDGIGNSARLAHGA VFGQY GIDRFYIGISAGAGFSSGSLSDGIGXKXRRRV LHYGIQA  
 orf1a XFDDGIGNSARLAHGA VFGQY GIGRFDIGISTGAGFSSGXLSDGIGGKIRRRV LHYGIQA  
 1260 1270 1280 1290 1300 1310  
 25 orf1.pep 890 900 910 920 930 940  
 RYRAGFGGFGIEPHIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSFKPAQHI  
 orf1a RYRAGFGGFGIEPYIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSFKPAQHX  
 1320 1330 1340 1350 1360 1370  
 30 orf1.pep 950 960 970 980 990 1000  
 SITPYLSLSYTDAA SGKVR TRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSLHAAA AKGP  
 orf1a SITPYXSLSYTDAA SGKVR TRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSXHAAA AKGP  
 1380 1390 1400 1410 1420 1430  
 40 orf1.pep 1010 1020  
 QLEAQHSAGIKLGYRWX  
 orf1a QLEAQHSAGIKLGYRWX  
 1440 1450

The complete length ORF1a nucleotide sequence <SEQ ID 651> is:

1 ATGAAACAA CCGACAAACG GACAACCGAA ACACACCGCA AAGCCCCGAA  
 45 51 AACCGGCCGC ATCCGCTTCT CGCCTGCTTA CTTAGCCATA TGCCTGTCTG  
 101 TCGGCATTCT TCCCAAGCT TGGCGGGGAC ACCTTATTT CGGCATCAAC  
 151 TACCAATACT ATCGCGACTT TGCCGAAAAT AAAGGCAAGT TTGCAGTCGG  
 201 GCGGAAAGAT ATTGAGTNT ACAACAAAA AGGGGAGTTG GTCGGCAAAT  
 251 CAATGACAAA AGCCCCGATG ATTGATTTT CTGTGGTGTC GCGTAACGGC  
 50 301 GTGGCGGCAT TGGTGGGCGA TCAATATATT GTGAGCGTGG CACATAACGG  
 351 CGGCTATAAC AACGTTGATT TTGGTGCGGA AGGAAGNAAT CCCGATCAGC  
 401 ACCGTTTTTC TTACCAATT GTGAAAAGAA ATAATTATAA GCCTGACAAT  
 451 TCACACCTTT ACAACGGCGA TTANCATATG CCGCGTTTGC ATAAATTTGT  
 501 CACAGATGCA GAACCTGTCG AAATGACGAG TGACATGAGG GGGAAATACCT  
 55 551 ATTCCGATAA AGAAAAATAT CCGAGCGTG TCCGCATCGG CTCAGGACAC  
 601 CACTATTGGC GTTATGATGA TGACAAACAC GGCATTTTAT CCTACTCCGG  
 651 CGCATGGTTA ATTGGCGGCA ATACACATAT GCAGGGTTGG GGAATAATG  
 701 GCGTANTTAG TTTGAGCGGC GATGTGCGCC ATGCCAACGA CTATGGCCCT  
 751 ATGCCGATTG CAGGTGCGGC AGGCGACAGC GGTCGCCAA TGTATTATTTA  
 801 TGACAAAACA AACAATAAAT GGCTGCTCAA CGGAGTTTTA CAAACCGGCT  
 851 ACCCTTATTC CGGCAGGAA AACGGTTTCC AGCTGATACG CAAAGATTGG  
 901 TTCTACGATG ACATTTACAG AGGCGATACA CATACCGTCT NTTTTGAACC  
 951 GCGCAGTAAC GGACATTTT CTTTACATC CAACAACAAC GGTACGGGTA  
 1001 CGGTAACAGA AACCAACGAA AAGGTNTCCA ATCCAAAGCT TAAAGTACAG  
 65 1051 ACAGTCCGAC TGTTTGACGA ATCTTTGAAT GAAACTGATA AAGAACCAGT  
 1101 TTACGCGGCA GGGGGTGTTA ATCAGTACCG TCCAAGGTTA AACACGGTG  
 1151 AAAACCTTTC TTTTATCGAT TACGGCAACG GCAAACTCAT CTTATCAAAC  
 1201 AACATCAACC AAGGCGCGGG CGGTTTGAT TTTGAAGGTG ATTTTACGGT

5

10

15

20

25

30

35

40

45

50

55

60

1251 CTCGCCTGAA AACAAACGAAA CGTGGCAAGG CGCGGGCGTT CATATCAGTG  
 1301 AAGACAGTAC CGTTACTTGG AAAGTAAACG GCGTGGCAAA CGACCCGCTG  
 1351 TCCAAAATCG GCAAAGGCAC GCTGCACGTT CAAGCCAAAG GGGAAAACCA  
 1401 AGGCTCGATC AGCGTGGGCG ACGGTACAGT CATTTTGGAT CAGCAGGCAG  
 1451 ACGATAAAGG CAAAAAACAA GCCTTTAGTG AAATCGGCTT GNTCAGCGGC  
 1501 AGGGGTACGG TGCAACTGAA TGCCGATAAT CAGTTCAACC CCGACAACCT  
 1551 CTATTTTCGGC TTTCGCGGCG GACGTTTGGA TTAAACGGG CATTCGCTTT  
 1601 CGTTCACCG TATTCAAAT ACCGATGAAG GGGCGATGAT TGNCNATCAT  
 1651 AATGCCACAA CAACATCCAC CGTTACCATT ACAGGGAATG AAAGTATTAC  
 1701 ACAACCGAGT GGTAAAGAATA TCAATAGACT TAATTACAGC AAAGAAATTG  
 1751 CCTACAACGG TTGGTTTGCG GAGAAAGATA CGACCAAAAC GAACGGGCGG  
 1801 CTCAACCTTG TTACACGCC CGCCGACAGG GACCGCACCC NGCTGCTTTC  
 1851 CGGCGGAACA AATTTAAACG GCAACATCAC GCAAAACAAAC GGCAAACCTGT  
 1901 TTTTCAGCGG CAGACCGACA CCGCACGCCT ACAATCATTT AGGAAGCGGG  
 1951 TGGTCAAAAA TGGAAGGTAT CCCACAAGGA GAAATCGTGT GGGACAACGA  
 2001 CTGGATCNAC CGCACGTTTA AAGCGGAAAA TTTCATATT CAGGGCGGGC  
 2051 AGGCGGTGAT TTCCCGCAAT GTTGCCAAAG TGAAGGCGA TTGNCATTTG  
 2101 AGCAATCACG CCCAAGCAGT TTTTGGTGTC GCACCGCATC AAAGCCATAC  
 2151 AATCTGTACA CGTTCGGACT GGACNNGTCT GACAAATTGT GTCGAANAAA  
 2201 NCATTACCGA CGATAAAGTG ATTGCTTCAT TGACTAAGAC NGACNTNAGC  
 2251 GGCANTGTNA GNCTNNCCNA TNACGNTNNT TNAAANCTCN CNGGGCGTGC  
 2301 NNCACNTAAN GGCAATCTTA GTGCAATGG CGATACAGT TATACAGTCA  
 2351 GCCACAACGC CACCCAAAAC GGCAACCTTA GCCTCGTGGG CAATGCCCCA  
 2401 GCAACATTTA ATCAAGCCAC ATTAAACGGC AACNCATCGG NTTCGGGCAA  
 2451 TGCTTCATTT AATCTAAGCA ACAACGCCGC AAAAAACGGC AGTCTGACGC  
 2501 TTTCCGACAA CGCTAAGGCA AACGTAAGCC ATTCCGCACT CAACGGCAAT  
 2551 GTCTCCCTAG CCGATAAGGC AGTATTCCAT TTTGAAAACA GCCGCTTTAC  
 2601 CGGACAACCTC AGCGGCAGCA AGGANACAGC ATTACACTTA AAAGACAGCG  
 2651 AATGGACGCT GCCGTACGGC ACGGAATTAG GCAATTTAAA CCTTGACAAC  
 2701 GCCACCATTA CACTCAATTC CGCTATCGC CACGATGCTG CAGGCGCGCA  
 2751 AACCGGCAGN GTGTCAGACA CGCCGCGCGC CGTTTCGCGC CGTTCCTAT  
 2801 TATCCGTTAC ACCGCCAATC TCGGTAGAAT CCCGTTTCAA CACGCTGACG  
 2851 GTAAACGGCA AATTGAACNG TCAAGGAACA TTCCGCTTTA TGTCGGAATC  
 2901 CTTTCGGCTAC CGAAGCGACA AATTGAAGCT GGCGGAAAGT TCCGAAGGNA  
 2951 CTTACACCTT GGCGGTCAAC AATACCGGCA ACGAACCCGT AAGCCTCGAT  
 3001 CAATTGACGG TAGTGGAAGG GAAAGACAAC AAACCGCTGT CCGAAAACCT  
 3051 TAATTTACCC TTGCAAAAACG AACACGTCGA TGCCGCGCGC TGGCGTTACC  
 3101 AACTCATCCG CAAAGACGGC GAGTTCGCC TGCATAATCC GGTCAAAGAA  
 3151 CAAGAGCTTT CCGCAAAACT CGGCAAGGCA GAAGCCAAA AACAGGCGGA  
 3201 AAAAGACAAC GCGCAAGGCG TTGACGCGCT GATTGCGGCC GGGCGCGATG  
 3251 CCGCCGAAAA GACAGAAAGC GTTGCCGAAC CGGCCCGGCN GGCAGGCGGG  
 3301 GAAAATGTCG GCATTATGCA GCGGAGGAA GAGAAAAAAC GGTGTCAGGC  
 3351 GGATAAAGAC AGCGCNTTGG CGAAACAGCG CGAAGCGGAA ACCCGGCCGG  
 3401 NTACCAACCG CTTCGCCCGC GCCCGCNGCG CCCGCCGGGA TTTGCCGCAA  
 3451 CCGCAGCCCC AACCGCAACC TCAACCCCAA CCGCAGCGCG ACCTGATNAG  
 3501 CCGTTATGCC AATAGCGGTT TGAGTGAATT TTCCGCCACG CTCAACAGCG  
 3551 TTTTCGCGCT ACAGGACGAA TTGGACCGCG TGTGTGCCGA AGACCGCCGC  
 3601 AACGCNGTTT GGACAAGCNG CATCCGNGAC ACCAAACACT ACCGTTTCGCA  
 3651 AGATTTCCGC GCCTACCGCC AACAAACCGA CCTGCGCCAA ATCGGTATGC  
 3701 AGAAAAACCT CGGCAGCGGG CGCGTCGGCA TCCTGTTTTT GCACAACCGG  
 3751 ACCGAAAAACA NCTTCGACGA CGGCATCGGC AACTCGGCAC GGCTTGCCCA  
 3801 CGGCGCCGTT TTCGGGCAAT ACGGCATCGG CAGGTTCGAC ATCGGCATCA  
 3851 GCACGGGCGC GGGTTTTAGC AGCGGCANTC TNCAGACGG CATCGGAGGC  
 3901 AAAATCCGCC GCCGCGTGCT GCATTACGGC ATTCAGGCAC GATACCGCGC  
 3951 CGGTTTCGGC GGATTTCGGC TCGAACCGTA CATCGGCGCA ACGCGCTATT  
 4001 TCGTCCAAAA AGCGGATTAC CGCTACGAAA ACGTCAATAT CGCCACCCCC  
 4051 GGTCTTGCGT TCAACCGNTA CCGNGCGGGC ATTAAGGCAG ATTATTCAAT  
 4101 CAAACCGGCG CAACACATNT CCATCACNCC TTATTTNAGC CTGTCTATA  
 4151 CCGATGCCGC TTCGGGCAAA GTCCGAACAC GCGTCAATAC CGCNGTATTG  
 4201 GCTCAGGATT TCGGCAAAAC CCGCAGTCG GAAATGGGCG TAAACGCCGA  
 4251 AATCAAAGT TACACGCTGT CCNTCCACGC TGCCGCGGCC AAAGNCCGC  
 4301 AACTGGAAGC GCAACACAGC GCGGCATCA AATTAGGCTA CCGCTGGTAA

This encodes a protein having amino acid sequence <SEQ ID 652>:

65

1 MKTTDKRTE THRKAPKTGR IRFSPAYLAI CLSFGILPQA WAGHTYFGIN  
 51 YQYYRDFEEN KGKFAVGAKD IEVYNKKGEL VGKSMTKAPM IDFSVVSRRNG  
 101 VAALVGDQYI VSAHNGGYN NVDFGAEGXN PDQHRFSYQI VKRNNYKPDN  
 151 SHPYNGDXHM PRLHKFVTD AEPVEMTSDMR GNTYSDEKEY PERVRIGSGH  
 201 HYWRYDDDKH GDLSYSGAWL IGGNTHMQGW GNNGVXSLSG DVRHANDYGP

5  
10  
15  
20  
25

```

251 MPIAGAAGDS GSPMFIYDKT NNKWLINGVL QTGYPYSGRE NGFQLIRKDW
301 FYDDIYRGDT HTVXFEPNSN GHFSFTSNNN GTGTVTETNE KVSNNPKLVQ
351 TVRLFDES LN ETDKEPVYAA GGVNQYRPRL NNGENLSFID YGNGKLILSN
401 NINQGAGGLY FEGDFTVSPE NNETWQAGAV HISEDSTVTW KVNGVANDRL
451 SKIGKGT LHV QAKGENQGSI SVGDGT VILD QQADDKGKKQ AFSEIGLXSG
501 RGTVQLNADN QFNPKLYFG FRGGRLDLNG HSLSFHRIQN TDEGAMIXXH
551 NATTTSTVTI TGNESITQPS GKNINRLNYS KEIAYNGWFG EKDTTKTNGR
601 LNLVYQPAAE DRTXLLSGGT NLNGNITQTN GKLFSSGRPT PHAYNHLGSG
651 WSKMEGIPQG EIVWDNDWIX RTFKAENFHI QGGQAVISRN VAKVEGDHXL
701 SNHAQAVFGV APHQSHITCT RSDWTGLTNC VEXXITDDKV IASLTKTDXS
751 GXVXLXXXXX XXLXGXAXLX GNLSANGDTR YTVSHNATQN GNLSLVGNAQ
801 ATFNQATLNG NXSXSGNASF NLSNNAQNG SLTSLDNAKA NVSHSALNGN
851 VSLADKAVFH FENSRTGQL SGSKXTALHL KDSEWTLPSG TELGNLNLDN
901 ATITLNSAYR HDAAGATGX VSDTPRRSR RSLSVTPPT SVESRFTLT
951 VNGKLNKQGT FRFMSELFY RSDKLKLAES SEGTYTLAVN NTGNEPVSLD
1001 QLTVEGKDN KPLSENINFT LQNEHVDAGA WRYQLIRKDG EFRLHNPVKE
1051 QELSDKLGA EAKKQAEKN AQSLDALIAA GRDAAEKTES VAEPARXAGG
1101 ENVGIMQAE EKKRVQADKD SALAKQREAE TREXTTAFPR ARKARRDLPO
1151 PQPQPQPQPQ PQRDLSRYA NSGLSEFSAT LNSVFAVQDE LDRVFAEDRR
1201 NAVWTSXIRX TKHYRSQDFR AYRQOTDLRQ IGMQKNLGS RVGILFSHNR
1251 TENXFDGIG NSARLAHGAV FGQYGIGRFD IGISTGAGFS SGXLSDGIGG
1301 KIRRRVLHYG IQARYRAGFG GFGIEPYIGA TRYFVQKADY RYENNVNIATP
1351 GLAFNRIRAG IKADYSFKPA QHXSITPYXS LSYTDAASGK VRTRVNTAVL
1401 AQDFGKTRSA EWGVNAEIKG FTLSXHAATA KGPQLEAQHS AGIKLGYRW*

```

25 A transmembrane region is underlined.

ORF1-1 shows 86.3% identity over a 1462aa overlap with ORF1a:

30  
35  
40  
45  
50  
55  
60  
65

```

          10      20      30      40      50      60
orfla.pep MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQAWAGHTYFGINYQYYRDFAE
|||||
orfl-1    MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQAWAGHTYFGINYQYYRDFAE
          10      20      30      40      50      60

          70      80      90     100     110     120
orfla.pep KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVSRRGVAALVGDDQYIVSVAHNGGYN
|||||
orfl-1    KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVSRRGVAALVGDDQYIVSVAHNGGYN
          70      80      90     100     110     120

          130     140     150     160     170     179
orfla.pep NVDFGAEGXNPDQHRFSYQIVKRNNYKPDNS-HPYNGDXHMPRLHKFVTDAPVEMTSDM
|||||
orfl-1    NVDFGAEGRNPDQHRFTYKIVKRNNYKAGTKGHPYGGDYHMPRLHKFVTDAPVEMTSYM
          130     140     150     160     170     180

          180     190     200     210     220     230
orfla.pep RGNTYSDEKYPERVIRIGSGHHYWRDDEKKGDL--SYSGA---WLIIGNTHMQGWGN
| | |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
orfl-1    DGRKYIDQNNYPDRVRIGAGRQYWRSEDEPNNRESSYHIASAYSWLVGNTFAQNGSGG
          190     200     210     220     230     240

          240     250     260     270     280     290
orfla.pep GVXSLSGD-VRHANDYGPMPIAGAAGDSGSPMFIYDKTNNKWLINGVLQTGYPYSGRENG
|:|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
orfl-1    GTVNLGSEKIKHS-PYGFLPTGGSGSGSPMFIYDAQKQKWLINGVLQTGNPYIGKSN
          250     260     270     280     290

          300     310     320     330     340     350
orfla.pep FQLIRKDWFYDDIYRGDTHTVXFEPNSNGHFSFTSNNNGTGTVTETNEKVSNNP-KLVQVT
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
orfl-1    FQLVRKDWFYDEIFAGDTHSVFYEPQNGKYSFNDDNNGTGKINAKHEHNSLPNRLKTRT
          300     310     320     330     340     350

          360     370     380     390     400     410
orfla.pep VRLFDES LN ETDKEPVY-AAGGVNQYRPRLNNGENLSFIDYNGKLILSNINQGAGGLY
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
orfl-1    VQLFNVSLSETAREPVYHAAGGVNSYRPRLNNGENISFIDEGKGELILTSNINQGAGGLY

```

		360	370	380	390	400	410	
			420	430	440	450	460	470
5	orf1a.pep		FEGDFTVSPENNETWQGAGVHISEDSTVTWKVNGVANDRLSKIGKGLHVQAKGENQGSII					
	orf1-1		FQGDFTVSPENNETWQGAGVHISEDSTVTWKVNGVANDRLSKIGKGLHVQAKGENQGSII					
		420	430	440	450	460	470	
10	orf1a.pep		480	490	500	510	520	530
	orf1-1		SVGDGTVILDQQADDKGGKQAFSEIGLVSGRGTVQLNADNQFNPDKLYFGFRGGRDLNG					
		480	490	500	510	520	530	
15	orf1a.pep		540	550	560	570	580	590
	orf1-1		HLSLFSHRIQNTDEGAMIXXHNATTSTVTITGNESITQPSGKNINRLNYSKEIAYNGWFG					
		540	550	560	570	580	590	
20	orf1a.pep		600	610	620	630	640	650
	orf1-1		EKDTTKTNGRLNLVYQPAEDRTXLLSGGTNLNGNITQTNGKLFFSGRPTPHAYNHLGSG					
25	orf1-1		EKDTTKTNGRLNLVYQPAEDRTLLSGGTNLNGNITQTNGKLFFSGRPTPHAYNHLNDH					
		600	610	620	630	640	650	
	orf1a.pep		660	670	680	690	700	710
	orf1-1		WSKMEGIPQGEIVWDNDWIXRTFKAENFHIQGGQAVISRNVAKEVGDXHLSNHAQAVFGV					
30	orf1-1		WSQKEGIPRGEIVWDNDWINRTFKAENFQIKGGQAVSRNVAKVKGDWHLNHAQAVFGV					
		660	670	680	690	700	710	
	orf1a.pep		720	730	740	750	760	770
35	orf1-1		APHQSHTICTRSDWTGLTNCVEKXITDDKVIASLTKTDISGNVDLADHAHLNLTGLATLN					
		720	730	740	750	760	770	
	orf1a.pep		780	790	800	810	820	830
40	orf1-1		GNLSANGDTRYTVSHNATQNGNLSLVGNAQATFNQATLNGNTSASGNASFNLSDHAVQNG					
		780	790	800	810	820	830	
45	orf1a.pep		840	850	860	870	880	890
	orf1-1		SLTSLDNAKANVSHSALNGNVSLADKAVFHFENSRTGQLSGSKXTALHLKDSEWTLPSG					
		840	850	860	870	880	890	
50	orf1a.pep		900	910	920	930	940	
	orf1-1		TELGNLNLDNATITLNSAYRHDAAGAQTGXVSDTPRRRSRRS---LLSVTPPTSVESRFN					
55	orf1-1		TELGNLNLDNATITLNSAYRHDAAGAQTGSATDAPRRRSRRSRRSLLSVTPPTSVESRFN					
		900	910	920	930	940	950	
	orf1a.pep		950	960	970	980	990	1000
	orf1-1		TLTVNGKLNQGTFRFMSELFGYRSDKLKLAESSEGTYTLAVNNTGNEPVSLEQLTVVEG					
60	orf1-1		TLTVNGKLNQGTFRFMSELFGYRSDKLKLAESSEGTYTLAVNNTGNEPASLEQLTVVEG					
		960	970	980	990	1000	1010	
	orf1a.pep		1010	1020	1030	1040	1050	1060
65	orf1-1		KDNKPLSENLFNLQNEHVDAGAWRYQLIRKDGFRHLNPNVKEQELSDKLGKAEAKKQAE					
		1020	1030	1040	1050	1060	1070	
70	orf1a.pep		1070	1080	1090	1100	1110	1120
			KDNAQSLDALIAAGRDAAEKTESVAEPARXAGGENVGIMQAEKKRVQADKDSALAKQR					

	orf1-1	 KDNAQSLDALIAAGRDAVEKTESVAEPAQAGGENVGIMQAEEEKKRVQADKDTALAKQR 1080 1090 1100 1110 1120 1130
5	orf1a.pep	1130 1140 1150 1160 1170 1180 EAETRPXTTAFPRARXARRDLPPQPQPQPQPQORDLXSRYANSGLSEFSATLNSVFAV 
10	orf1-1	EAETRPATTAFPRARRARDLPQLQPQPQPQP--QRDLISRYANSGLSEFSATLNSVFAV 1140 1150 1160 1170 1180 1190
	orf1a.pep	1190 1200 1210 1220 1230 1240 QDELDRVFAEDRRNAVWTSXIRXTKHYRSQDFRAYRQQTDLRQIGMQKNLGSGRVGILFS 
15	orf1-1	QDELDRVFAEDRRNAVWTSGIRDTKHYRSQDFRAYRQQTDLRQIGMQKNLGSGRVGILFS 1200 1210 1220 1230 1240 1250
	orf1a.pep	1250 1260 1270 1280 1290 1300 HNRTENXFDDGIGNSARLAHGAVFGQYIGRFDIGISTGAGFSSGXLSDGIGGKIRRRVL 
20	orf1-1	HNRTENTFDDGIGNSARLAHGAVFGQYIGRFDIGISTGAGFSSGXLSDGIGGKIRRRVL 1260 1270 1280 1290 1300 1310
	orf1a.pep	1310 1320 1330 1340 1350 1360 HYGIQARYRAGFGGFGIEPIYIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSF 
25	orf1-1	HYGIQARYRAGFGGFGIEPIYIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSF 1320 1330 1340 1350 1360 1370
	orf1a.pep	1370 1380 1390 1400 1410 1420 KPAQHXSITPYXSLSYTDAASGKVTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSXHA 
30	orf1-1	KPAQHISITPYLSLSYTDAAASGKVTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSLHA 1380 1390 1400 1410 1420 1430
	orf1a.pep	1430 1440 1450 AAAKGPQLEAQHSAGIKLGYRWX 
35	orf1-1	AAAKGPQLEAQHSAGIKLGYRWX 1440 1450

Homology with adhesion and penetration protein hap precursor of *H. influenzae* (accession number P45387)

Amino acids 23-423 of ORF1 show 59% aa identity with hap protein in 450aa overlap:

	orf1	23	FXAAYLAICLSFGILPQAWAGHTYFGINYQYYRDAENKGGKFAVGAKDIEVYNKKGELVG	82
45	hap	6	FRLNFLTACVSLGIASQAWAGHTYFGIDYQYYRDAENKGGKFTVGAKNIEVYNKEGQLVG	65
	orf1	83	KSMTKAPMIDFSVSRNGVAALVGQYIVSVAHNGGYNVDFGAEGXNIXDQXRTYKIV	142
50	hap	66	TSMTKAPMIDFSVSRNGVAALVGQYIVSVAHNGGYNVDFGAEGN DQ R TY+IV	124
	orf1	143	KRNNYKAGTKGHPYGGDYHMPRLHKXVTDAPVEMTSYMDGRKYIDQNNYPDRVRIGAGR	202
	hap	125	KRNNYQAWERKHPYDGDYHMPRLHKFVTEAEPVGMTTNMDGKVYADRENYPERVRIGSGR	184
55	orf1	203	QYWRSDDEPNRESSYHIA-----	222
	hap	185	QYWRD+DE N SSY+++	
	orf1	223	QYWRD+DE N SSY+++	
60	hap	245	GSKGDSGSPMFIYDAKKQWLINAVLQTHGPFGRGNGFQLIREEFYNEVLAVDTPSVF	304
	orf1	278	--YEPRQNGKYSFNDDNNGTGKIN-AKHEHNSLPNRLKTRTVQLFNVSLSSETAREPVYHA	334
65	hap	305	QRYIPPINGHYSFVSNNDGTGKLTLTRPSKDGSKAKSEVGTVKLFNPSLNQTAKEHV-KA	363



5 orf1 335 AGGVNSYRPRLNNGENISFIDEGKGELILTSNINQGAGGLYFQGDFTV-SPENNETWQGA 393  
 A G N Y+PR+ G+NI D+GKG L + +NINQGAGGLYF+G+F V +NN TWQGA  
 hap 364 AAGYNIYQPRMEYGKNIYLGQDGKGTLTNIENNINQGAGGLYFEGNFVVKGKQNNITWQGA 423

orf1 394 GVHISEDSTVTWKVNGVANDRLSKIGKGT 423  
 GV I +D+TV WKV+ NDRLSKIG GTL  
 hap 424 GVSIGQDATVEWKVHNPENDRLSKIGIGTL 453

Amino acids 715-1011 of ORF1 show 50% aa identity with hap protein in 258aa overlap:

10 Orf1 41 DTRYTVSHNATQ-NGNXSLVXNAQATFNQ-ATLNGNTSASGNASENLSDHAVQNGSLTLS 98  
 DT+ S TQ NG+ +L NA + A LNGN + ++ F LS++A Q G++ LS  
 hap 733 DTKVINSIPITQINGSINLTNNATVNIHGLAKLNGNVTLDHSQFTLSNNATQTGNIKLS 792

orf1 99 GNAKANVSHSALNGNVSLADKAVHFESSRFTGQISGGKDTALHLKDSEWTLPSGXELGN 158  
 +A A V+++ LNGNV L D A F ++S F QI G KDT + L+++ WT+PS L N  
 15 hap 793 NHANATVNNATLNGNVHLTDSAQFSLKNSHFWHQIQGDKDTTVTLENATWTMPSTDTLQN 852

orf1 159 LNLDNATITLNSAYRHAAGAQTGSATDAPXXXXXXXXXXLLXVTPPTSVESRFTLTVN 218  
 L L+N+T+TLNSAY + S+ +AP L T PTS E RFTLTVN  
 20 hap 853 LTLNNSTVTLNSAY-----SASSNNAPRHRS-----LETETTP TSAEHRFTLTVN 899

orf1 219 GKLNQGTFRFMSELFYRSDKLKLAESSEGTTLAVNNTGNEPASLEQLTVVEGKDNKP 278  
 GKL+GQGTFF S LFGY+SDKLKL+ +EG YTL+V NTG EP +LEQLT++E DNKP  
 hap 900 GKLSGQGTFFQFTSSLFYKSDKLKLSNDAEGDYTLVRNTGKEPVTLEQLTLIESLDNKP 959

25 orf1 279 LSENLFNFTLQNEHVDAGA 296  
 LS+ L FTL+N+HVDAGA  
 hap 960 LSDKLKFTLENDHVDAGA 977

Amino acids 1192-1450 of ORF1 show 41% aa identity with hap protein in 259aa overlap:

30 Orf1 1 LDRVFAEDRRNAVWTSIGIRDTKHYRSQDFRAYRQQTDLRQIGMQKNLGSGRVGILFHSNR 60  
 LDR+F + ++AVWT+ +D + Y S FRAY+Q+T+LRQIG+QK L +GR+G +FSH+R  
 hap 1135 LDRLFVDQAQSAVWNTNIAQDKRRYDSDAFRAYQQKTNLRLQIGVQKALANGRIGAVFHSR 1194

orf1 61 TENTFDDGIGNSARLAHGA VFGQY GIDRFYXXXXXXXXXXXXXXXXXIGKXRRRVLHYG 120  
 ++NTFD+ + N A L + F QY K R+ ++YG  
 35 hap 1195 SDNTFDEQVKNHATLTMMSGFAQYQWGDQLFGVNVGTGISASKMAEEQSRKIHRKAINYG 1254

orf1 121 IQARYRAGFGGFGIEPHIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSEKPA 180  
 + A Y+ G GI+P+ G RYF+++ +Y+ E V + TP LAFNRY AGI+ DY+F P  
 40 hap 1255 VNASYQFRLGLQGIQPYFGVNRIFYERENYQSEEVVRKTPSLAFNRYNAGIRVDYFTPT 1314

orf1 181 QHISITPYLSLSYTDASGKVRTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSLHAAA 240  
 +IS+ PY ++Y D ++ V+T VN VL Q FG+ E G+ AEI F +S + +  
 hap 1315 DNISVKPYFFVNYVDVSNANVQTTVNLTVLQPFGRYWQKEVGLKAEILHFQISAFISKS 1374

45 orf1 241 KGPQLEAQHSAGIKLGYRW 259  
 +G QL Q + G+KLG YRW  
 hap 1375 QGSQLGKQONVGVKLG YRW 1393

#### Homology with a predicted ORF from *N.gonorrhoeae*

50 The blocks of ORF1 show 83.5%, 88.3%, and 97.7% identities in 467, 298, and 259 aa overlap, respectively with a predicted ORF (ORF1ng) from *N.gonorrhoeae*:

orf1.pep MKTTDKRTTETHRKAPKTGRIRFXAAYLAICLSFGILPQAWAGHTYFGINYQYYRDFAE 60  
 |||||  
 55 orf1ng MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQARAGHTYFGINYQYYRDFAE 60

orf1.pep KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVSRRNGVAALVGVQYIVSVAHNGGYN 120  
 |||||  
 orf1ng KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVSRRNGVAALAGDQYIVSVAHNGGYN 120

60 orf1.pep NVDFGAEGXNIXDQXRTYKIVKRNNYKAGTKGHPYGGDYHMPRLHKKVTDAPVEMTSY 180  
 |||||  
 orf1ng NVDFGAEGSN-PDQHRFSYQIVKRNNYKAGTNGHPYGGDYHMPRLHKKVTDAPVEMTSY 179

	orf1.pep	MDGRKYIDQNNYPDRVRIGAGRQYWRSEDEPNNRESSYHIAS-----	223
	orf1ng	MDGWKYADLNKYPDRVRIGAGRQYWRSEDEPNNRESSYHIASAYSWLVGGNTFAQNGSG	239
5	orf1.pep	-----GSEPMFIYDAQKQKWLINGVLOTGNPYIGKSNG	255
	orf1ng	GGTVNLGSEKIKHSPYGFLPTGGSFGDSGSEPMFIYDAQKQKWLINGVLOTGNPYIGKSNG	289
10	orf1.pep	<u>FQLVRKDWFYDEIFAGDTHSVFYEPHQNGKYSFNDNNGTGKINAKHEHNSLPNRLKTRT</u>	315
	orf1ng	<u>FQLVRKDWFYDEIFAGDTHSVFYEPHQNGKYFFNDNNGAGKIDAKHKHYSLPYRLKTRT</u>	359
15	orf1.pep	VQLFNVSLSETAREPVYHAAGGVNSYRPLNNGENISFIDEGKGELILTSNINQGAGGLY	375
	orf1ng	VQLFNVSLSETAREPVYHAAGGVNSYRPLNNGENISFIDKKGELILTSNINQGAGGLY	
	orf1.pep	FQGDFTVSPENNETWQAGVHISEDSTVTWKVNGVANDRLSKIGKGT	422
20	orf1ng	FEKNFTVSPKNNETWQAGVHISDGSTVTWKVNGVANDRLSKIGKGTLLVQAKGENQGSV	479
	orf1.pep	// DKVTASLTKTDISGNVOLDHAHLNLTGLA	744
	orf1ng	FGVAPHQSHTICTRSDWTGLTSCTEKITITDDKVIASLTKTDVRGNVSLADHAHLNLTGLA	774
25	orf1.pep	TLNGNLSANGDTR-YTVSHNATQNGXSLVXNAQATFNQATLNGNTSASGNASFNLSDBA	803
	orf1ng	TFNGNL-VQAEFTRIRLRANATQNGNLSLVGNAQATFNQATLNGNTSASDNASFNLSNNA	833
30	orf1.pep	VQNGSLTSLGNANVSHSALNGVSLADKAVFHFESSRFTQISGGKDTALHLKDSEWT	863
	orf1ng	VQNGSLTSLSDNANVSHSALNGVSLADKAVFHFENSRTGKISGGKDTALHLKDSEWT	893
35	orf1.pep	LPSGXLGNLNLNATITLNSAYRHAAGAQTGSATDAPRRSRSSRSLXVTPPTSVE	923
	orf1ng	LPSGTELGNLNLNATITLNSAYRHAAGAQTGSAADAPRRSRSSR---LLSVTPPTSVE	950
	orf1.pep	SRFNTLTVNGKLNQGTFRFMSELFYRSDKLKLAESSEGTYTLAVNNTGNEPASLEQLT	983
40	orf1ng	SRFNTLTVNGKLNQGTFRFMSELFYRSGKLKLAESSEGTYTLAVNNTGNEPVSLEQLT	1010
	orf1.pep	VVEGKDNKPLSENLFNLQNEHVDAGAW	1011
45	orf1ng	VVEGKDNTPLESENLFNLQNEHVDAGAWRYQLIRKDGFRHLNHPVKEQELSDKLKAGET	1070
	orf1.pep	// LDRVFAEDRRNAVVTSGIRDTKHYRSQDFR	1211
	orf1ng	PQRDLISRYANSGLSEFSATLNSVFAVQDELDRVFAEDRRNAVVTSGIRDTKHYRSQDFR	1239
50	orf1.pep	AYRQQTDLRQIGMQKNLGSGRVGILFSHNRTENTFDDGIGNSARLAHGAVFGQYIGDRFY	1271
	orf1ng	AYRQQTDLRQIGMQKNLGSGRVGILFSHNRTGNTFDDGIGNSARLAHGAVFGQYIGGRFD	1299
55	orf1.pep	IGISAGAGFSSGSLSDGIGKXRRRVLYHYGIQARYRAGFGGFGIEPHIGATRYFVQKADY	1331
	orf1ng	IGISAGAGFSSGSLSDGIRGKIRRRVLYHYGIQARYRAGFGGFGIEPHIGATRYFVQKADY	1359
	orf1.pep	RYENVNIATPGLAFNRYRAGIKADYSFKPAQHISITPYLSLSYTDAAAGKVRTRVNTAVL	1391
60	orf1ng	RYENVNIATPGLAFNRYRAGIKADYSFKPAQHISITPYLSLSYTDAAAGKVRTRVNTAVL	1419
	orf1.pep	AQDFGKTRSAEWGVNAEIKGFTLSLHAAAAGPQLEAQHSAGIKLGYRW 1440	
	orf1ng	AQDFGKTRSAEWGVNAEIKGFTLSLHAAAAGPQLEAQHSAGIKLGYRW 1468	

65 The complete length ORF1ng nucleotide sequence was identified <SEQ ID 653>:

```

1  ATGAAAACAA CCGACAAACG GACAACCGAA ACACACCGCA AAGCCCCTAA
51  AACCGGCCGC ATCCGCTTCT CGCCCGCTTA CTTAGCCATA TGCCTGTCGT
101 TCGGCATTCT GCCCCAAGCC CGGGCGGGAC AACTTATTT CGGCATCAAC

```

5

10

15

20

25

30

35

40

45

50

55

60

65

70

151 TACCAATACT ATCGCGACTT TGCCGAAAAT AAAGGCAAGT TTGCAGTCGG  
201 GGCGAAAGAT ATTGAGGTTT ACAACAAAA AGGGGAGTTG GTCCGGCAAT  
251 CGATGACGAA AGCCCCGATG ATTGATTTT CTGTGGTATC GCGTAACGGC  
301 GTGGCGGCAT TGGCGGGCGA TCAATATATT GTGAGCGTGG CACATAACGG  
351 CGGCTATAAC AATGTTGATT TTGGTGCGBA GGAAGCAAT CCCGATCAGC  
401 ACCGCTTTTC TTACCAAATT GTGAAAAGAA ATAATTATAA AGCAGGGACT  
451 AACGGCCATC CTTATGGCGG CGATTATCAT ATGCCGCGTT TGCACAAAT  
501 TGTCACAGAT GCAGAACCCTG TTGAGATGAC CAGTTATATG GATGGGTGGA  
551 AATACGCTGA TTAAATAAA TACCCTGATC GTGTTTCAAT CGGAGCAGGC  
601 AGACAATATT GGCGGTCTGA TGAAGACGAA CCAATAAACC GCGAAAGTTC  
651 ATATCATATT GCAAGCGCAT ATTCTTGGCT CGTCGGTGGC AATACCTTTG  
701 CACAAAATGG ATCAGGTGGT GGCACAGTCA ACTTAGGTAG CGAAAAAATT  
751 AAACATAGCC CATATGGTTT TTACCAACA GGAGGCTCAT TTGGCGACAG  
801 TGGCTCACCA ATGTTTATCT ATGATGCCCA AAAGCAAAAG TGGTTAATTA  
851 ATGGGGTATT GCAAAACAGC AACCCCTATA TAGGAAAAAG CAATGGCTTC  
901 CAGCTAGTTC GTAAAGATTG GTTCTATGAT GAAATCTTTG CTGGAGATAC  
951 CCATTCAGTA TTCTACGAAC CACATCAAAA TGGGAAATAC TTTTTTAACG  
1001 ACAATAATAA TGGCGCAGGA AAAATCGATG CCAACATAA ACCTATTCTT  
1051 CTACCTTATA GATTAATAAAC ACGAACCCTT CAATTGTTTA ATGTTTCTTT  
1101 ATCCGAGACA GCAAGAGAAC CTGTTTATCA TGCTGCAGGT GGGGTCAACA  
1151 GTTATCGACC CAGACTGAAT AATGGAGAAA ATATTTCTCT TATTGACAAA  
1201 GGAAAAGGTG AATTGATACT TACCAGCAAC ATCAACCAAG GCGCGGGCGG  
1251 TTTGTATTTT GAGGGTAATT TTACGGTCTC GCCTAAAAAC AACGAAACGT  
1301 GGCAAGGCGC GGGCGTTTAT ATCAGTGATG GCAGTACCGT TACTTGGAAG  
1351 GTAAACGGCG TGGCAACGCA CCGCTGTCC AAAATCGGCA AAGGCACGCT  
1401 GCTGGTTCAA GCCAAAGGGG AAAACCAAGG CTCGGTCAGC GTGGGCGAGC  
1451 GTAAAGTCAT CTTAGATCAG CAGGCGGAGC ATCAAGGCAA AAAACAAGCC  
1501 TTTAGTGAAA TCGGCTTGGT CAGCGGCAGG GGGACGGTGC AACTGAATGC  
1551 CGATAATCAG TTCAACCCCG ACAACTCTA TTTGCGCTTT CGCGGCGGAC  
1601 GTTTGGATTT GAACGGGCAT TCGCTTTCGT TCCACCGCAT TCAAAATACC  
1651 GATGAAGGGG CGATGATTGT CAACCACAT CAAGACAAAG AATCCACCGT  
1701 TACCATTACA GGCAATAAAG ATATTACTAC AACCGGCAAT AACACAACCT  
1751 TGGATAGCAA AAAAGAAATT GCCTACAACG GTTGGTTTGG CGAGAAAGAT  
1801 GCAACCAAAA CGAACGGGCG GCTCAATCTG AATTACCAAC CGGAAGAAGC  
1851 GGATCGCACT TTTACTGCTT CCGCGGAAC AAATTTAAAC GGCAATATCA  
1901 CGCAACAAA CGGCAAACTG TTTTCAGCG GCAGACCGAC ACCGACCGC  
1951 TACAATCATT TAGGAAGCGG GTGGTCAAAA ATGGAAGGTA TCCACAAGG  
2001 AGAAATCGTG TGGGACAACG ATTGGATCGA CCGCACATT AAAGCGGAAA  
2051 ACTTCCATAT TCAGGGCGGA CAAGCGGTGG TTTCCCGCAA TGTTCGCAAA  
2101 GTGGAAGGCG ATTGGCATT TTAGCAATCAC GCCCAAGCAG TTTTCGGTGT  
2151 CGCACCGCAT CAAAGCCACA CAATCTGTAC ACGTTCGGAC TGGACGGGTC  
2201 TGACAAGTTG TACCGAAAAA ACCATTACCG ACGATAAAGT GATTGCTTCA  
2251 TTGAGCAAGA CCGACATCAG AGGCAATGTC AGCCTTGCCG ATCAGCTCA  
2301 TTAAATCTC ACAGGACTTG CCACACTCAA CGGCAATCTT AGTGCAGGCG  
2351 GAGACACGCA CTATACGGTT ACGCGCAACG CCACCAAAA CGGCAACCTC  
2401 AGCCTCGTGG GCAATGCCCA AGCAACATTT AATCAAGCCA CATTAAACGG  
2451 CAACACATCG GCTTCGGACA ATGCTTCATT TAATCTAAGC AACACGCGC  
2501 TACAAAACGG CAGTCTGACG CTTTCCGACA ACGCTAAGGC AACGTAAGC  
2551 CATTCCGCAC TCAACGGCAA TGTCTCCCTA GCCGATAAGG CAGTATTCCA  
2601 TTTTGAATAC AGCCGCTTTA CCGGAAAAAT CAGCGGCGGC AAGGATACGG  
2651 CATTACACTT AAAAGACAGC GAATGGACGC TGCCGTGCGG CACGGAATTA  
2701 GGCAATTTAA ACCTTGACAA CGCCACCATT AACTCAATT CCGCTATCG  
2751 ACACGATGCG GCAGGCGCGC AAACCGGCAG TGCGGCAGAT GCGCCGCGC  
2801 GCCGTTTCGCG CCGTTCCCTA TTATCCGTTA CGCCGCCAAC TTCGGCAGAA  
2851 TCCCGTTTCA ACACGCTGAC GGTAAACGGC AAATTGAACG GTCAGGGAAC  
2901 ATTCCGCTTT ATGTCGGAAC TCTTCGGCTA CCGCAGCGGC AAATTGAAGC  
2951 TGGCGGAAAG TTCCGAAGGC ACTTACACCT TGGCTGTCAA CAATACCGGC  
3001 AACGAACCCG TAAGTCTCGA GCAATTGACG GTAGTGGAAG GAAAAGACAA  
3051 CACACCGCTG TCCGAAATC TTAATTTAC CCTGCAaaAc gaacacgtcg  
3101 atgcccgcgc atggCGTTAT CAGCTTATCC gcaaaagacgG CGAGTTCCGc  
3151 CTGCATAATC CGGTCAAAGA ACAAGAGCTT TCCGACAAC TCGGCAAGGc  
3201 gggagaaACA GAggcccgcT TGACGGCAAA ACAGGCacaA CTGCGGCCA  
3251 AAcaacaggg ggaAAAAGAC AACgcgcaaa gccttgAcg gctgattgcy  
3301 gCggggcgca atgccaccga AAAGGCAgaa agtgttgccg aaccgGCCGc  
3351 GCAGGCAAGC GGGGAAAAtg ccgGCATTAT GCAGGCGGAG GAAGAGAAAA  
3401 AACGGGTGCA GCGGATAAA GACACCGCCT TGGCGAAACA GCGCGAAGCG  
3451 GAAACCCGCG CGGTACCAC CGCCTTCCCC CGCGCCCGCC GCGCCCGCGC  
3501 GATTTTGCCG CAACCGCAGC CCAACCGCA ACCCAACCG CAGCGCGACC  
3551 TGATCAGCCG TTATGCCAAT AGCGGTTTGA GTGAATTTTC CGCCACGCTC  
3601 AACAGCGTTT TCGCCGTACA GGACGAATTG GACCGCGTGT TTGCCGAAGA

5  
10  
15

```

3651 CCGCCGCAAC GCCGTTTGGG CAAGCGGCAT CCGGGACACC AAACACTACC
3701 GTTCGCAAGA TTCCGCGGCC TACCGCCAAC AAACCGACCT GCGCCAAATC
3751 GGTATGCAGA AAAACCTCGG CAGCGGGCGC GTCGGCATCC TGTTTTCGCA
3801 CAACCGGACC GGAAACACCT TCGACGACGG CATCGGCAAC TCGGCACGGC
3851 TTGCCCACGG TGCCGTTTTC GGGCAATACG GCATCGGCAG GTTCGACATC
3901 GGCATCAGCG CGGGCGCGGG TTTAGTAGC GGCAGCCTTT CAGACGGCAT
3951 CAGAGGCAAA ATCCGCCGCC GCGTGCTGCA TTACGGCATT CAGGCAAGAT
4001 ACCGCGCAGG TTTCGGCGGA TTCGGCATCG AACCGCACAT CGGCGCAACG
4051 CGCTATTTTC TCCAAAAGC GGATTACCGA TACGAAAACG TCAATATCGC
4101 CACCCCGGGC CTTCGATTCA ACCGCTACCG CGCGGGCATT AAGGCAGATT
4151 ATTCATTCAA ACCGGCGCAA CACATTTCCA TCACGCCTTA TTTAGCCTG
4201 TCCTATACCG ATGCCGCTTC CGGCAAAGTC CGAACGCGCG TCAATACCGC
4251 CGTATTGGCG CAGGATTTTC GCAAAACCG CAGTGCGGAA TGGGGCGTAA
4301 ACGCCGAAAT CAAAGGTTTC ACCTGTCTCC TCCACGCTGC CGCCGCCAAG
4351 GGGCCGCAAT TGAAGCGCA GCACAGCGCG GGCATCAAT TAGGCTACCG
4401 CTGTAA

```

This is predicted to encode a protein having amino acid sequence <SEQ ID 654>:

20  
25  
30  
35  
40  
45

```

1 MKTTDKRTE THRKAPKTGR IRFSPAYLAI CLSFGILPQA RAGHTYFGIN
51 YQYYRDAFEN KGKFAVGAKD IEVYNKKGEL VGKSMTKAPM IDFSVVSRRNG
101 VAALAGDQYI VSAHNGGYN NVDFGAEGSN PDQHRFSYQI VKRNNYKAGT
151 NGHPYGGDYH MPRLHKFVTD AEPVEMTSYM DGWKYADLNK YPDRVRIGAG
201 RQYWRSEDE PNNRESSYHI ASAYSWLVGG NTFQNGSGG GTVNLGSEKI
251 KHSYPYGFLT GGSFGDSGSP MFIYDAQKQK WLINGVLQTG NPYIGKSNGF
301 QLVRKDWFYD EIFAGDTHSV FYEPHQNGKY FFDNNNGAG KIDAKHKHYS
351 LPYRLKTRTV QLFNVSLSET AREPVYHAAG GVNSYRPRLN NGENISFIDK
401 GKGLILTSN INQAGGLYF EGNFTVSPKN NETWQAGAVH ISDGSTVTWK
451 VNGVANDRLS KIGKGTLLVQ AKGENQGSVS VGDGKVILDQ QADDQGGKQA
501 FSEIGLVSGR GTVQLNADNQ FNPDKLYFGF RGRLDLNGH SLSFHRIQNT
551 DEGAMIVNHN QDKESTVTIT GNKDITTTGN NNNLDSKKEI AYNWGFGEKD
601 ATKTNGLNL NYPPEEADRT LLLSGGTNLN GNITQTNGKL FFSGRPTPHA
651 YNHLGSGWSK MEGIPQGEIV WDNWDIDRTF KAENFHIQGG QAVVSRNVAK
701 VEGDWHLNSH AQAVFGVAPH QSHTICTRSD WTGLTSCTEK TITDDKVIAS
751 LSKTDVGRNV SLADHAHLNL TGLATFNGNL VQAEIRTRL RANATQNGNL
801 SLVGNAQATF NQATLNGNTS ASDNASFNLS NNAVQNGSLT LSDNAKANVS
851 HSALENGVSL ADKAVFHFEN SRFTGKISGG KDTALHLKDS EWTLPSTGTEL
901 GNLNLDNATI TLNSAYRHDA AGAQTGSAAD APRRRSRSL LSVTPPTSAE
951 SRFNTLTVNG KLNQGGTFRF MSELFYGRSG KLKLAESSEG TYTLAVNNTG
1001 NEPVSLQLT VVEGKDNTPL SENLFTLQN EHVDAWARY QLIRKDGFR
1051 LHNVPKEQEL SDKLGKAGET EAALTAKQAO LAAKQQAED NAQSLDALIA
1101 AGRNATEKAE SVAEPARQAG GENAGIMQAE EEKRVQADK D'ALAKQREA
1151 ETRPATTAF RARRARRDLP QPQPQPQPQ QRDILISRYN SGLSEFSATL
1201 NSVFAVQDEL DRVFAEDRRN AVWTSGIRD KHYRSQDFRA YRQOTDLRQI
1251 GMQKNLGSGR VGILFSHNRT GNTFDDGIGN SARLAHGAVF GQYIGIRFDI
1301 GISAGAGFSS GSLSDGIRGK IRRRVLHYGI QARYRAGFGG FGIEPHIGAT
1351 RYFVQKADYR YENVNIATPG LAFNRYRAGI KADYSFKPAQ HISITPYLSL
1401 SYTDAASGV RTRVNTAVLA QDFGKTRSAE WGVNAEIKGF TSLHAAAAK
1451 GPQLEAQHSA GIKLGYRW*

```

Underlined and double-underlined sequences represent the active site of a serine protease (trypsin family) and an ATP/GTP-binding site motif A (P-loop).

50 ORF1-1 and ORF1ng show 93.7% identity in 1471 aa overlap:

55  
60

```

          10      20      30      40      50      60
orfl-1.pep MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQAWAGHTYFGINYQYYRDAFEN
          10      20      30      40      50      60
orflng-1   MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQARAGHTYFGINYQYYRDAFEN

          70      80      90      100     110     120
orfl-1.pep KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVSRRNGVAALVGDQYIVSAHNGGYN
          70      80      90      100     110     120
orflng-1   KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVSRRNGVAALAGDQYIVSAHNGGYN

```



		790	800	810	820	830	840
		850	860	870	880	890	900
5	orf1-1.pep	LSGNAKANVSHSALNGVSLADKAVFHFESSRFTGQISGGKDTALHLKDSEWTLPSGTTEL					
	orf1ng-1	LSDNAKANVSHSALNGVSLADKAVFHFENSRTGKISGGKDTALHLKDSEWTLPSGTTEL					
		850	860	870	880	890	900
		910	920	930	940	950	960
10	orf1-1.pep	GNLNLDNATITLNSAYRHDAAAGQTSATDAPRRRSRRSRLSVTPPTSVERFNTLT					
	orf1ng-1	GNLNLDNATITLNSAYRHDAAAGQTSADAPRRRSR---RSLSVTPPTSVERFNTLT					
		910	920	930	940	950	
		970	980	990	1000	1010	1020
15	orf1-1.pep	VNGKLNQQTFRFMSELFGYRSDKLKLAESSEGTYYTLAVNNTGNEPASLEQLTVVEGKDN					
	orf1ng-1	VNGKLNQQTFRFMSELFGYRSGKLKLAESSEGTYYTLAVNNTGNEPVSLEQLTVVEGKDN					
		960	970	980	990	1000	1010
20		1030	1040	1050	1060	1070	
	orf1-1.pep	KPLSENLFNTLQNEHVDAGAWRYQLIRKDGFEFLHNPVKEQELSDKLKGA-----					
	orf1ng-1	TPLSENLFNTLQNEHVDAGAWRYQLIRKDGFEFLHNPVKEQELSDKLKGAETEALTA					
25		1020	1030	1040	1050	1060	1070
		1080	1090	1100	1110	1120	
	orf1-1.pep	----EAKKQAEKDNAQSLDALIAAGRAVEKTESVAEPARQAGGENVGMQAEKEKRVQ					
	orf1ng-1	QAQLAAKQAEKDNAQSLDALIAAGRNATEKAESVAEPARQAGGENAGIMQAEKEKRVQ					
30		1080	1090	1100	1110	1120	1130
		1130	1140	1150	1160	1170	1180
	orf1-1.pep	ADKDTALAKQREAEATPATTAFFPRARRARDLPQLQPQPQPQORDLISRYANSGLSEFS					
	orf1ng-1	ADKDTALAKQREAEATPATTAFFPRARRARDLPQPQPQPQPQORDLISRYANSGLSEFS					
35		1140	1150	1160	1170	1180	1190
		1190	1200	1210	1220	1230	1240
	orf1-1.pep	ATLNSVFAVQDELDRVFAEDRRNAVWTSIGIRDTKHYRSQDFRAYRQQTDLRQIGMQKNLG					
	orf1ng-1	ATLNSVFAVQDELDRVFAEDRRNAVWTSIGIRDTKHYRSQDFRAYRQQTDLRQIGMQKNLG					
40		1200	1210	1220	1230	1240	1250
		1250	1260	1270	1280	1290	1300
	orf1-1.pep	SGRVGILFSHNRTENTFDDGIGNSARLAHGAVFGQYIGIDRFYIGISAGAGFSSGSLSDGI					
	orf1ng-1	SGRVGILFSHNRTGNTFDDGIGNSARLAHGAVFGQYIGIRFDIGISAGAGFSSGSLSDGI					
45		1260	1270	1280	1290	1300	1310
		1310	1320	1330	1340	1350	1360
	orf1-1.pep	GGKIRRRVLHYGIQARYRAGFGGFGIEPHIGATRYFVQKADYRYENVNIATPGLAFNRYR					
	orf1ng-1	RGKIRRRVLHYGIQARYRAGFGGFGIEPHIGATRYFVQKADYRYENVNIATPGLAFNRYR					
50		1320	1330	1340	1350	1360	1370
		1370	1380	1390	1400	1410	1420
	orf1-1.pep	AGIKADYSFKPAQHISITPYLSLSYTDAAAGKVRTRVNTAVLAQDFGKTRSAEWGVNAEI					
	orf1ng-1	AGIKADYSFKPAQHISITPYLSLSYTDAAAGKVRTRVNTAVLAQDFGKTRSAEWGVNAEI					
55		1380	1390	1400	1410	1420	1430
		1430	1440	1450			
	orf1-1.pep	KGFTLSLHAAAAGPQLEAQSAGIKLGYRWX					
	orf1ng-1	KGFTLSLHAAAAGPQLEAQSAGIKLGYRWX					
60		1440	1450	1460			

In addition, ORF1ng shows 55.7% identity with hap protein (P45387) over a 1455aa overlap:

SCORES Init1: 1104 Initn: 4632 Opt: 2680  
Smith-Waterman score: 5165; 55.7% identity in 1455 aa overlap

5 orflng-1.pep MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQARAGHTYFGINYYRDFAE  
p45387 MKKTVFRLNFLTACISLGIIVSQAWAGHTYFGIDYQYYRDFAE  
10 orflng-1.pep KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVSRNGVAALAGDQYIVSVAHNGGYN  
p45387 KGKFTVGAQNIKVYNKQQLVGTSMTKAPMIDFSVSRNGVAALVENQYIVSVAHNVGYT  
15 orflng-1.pep NVDFGAEGSNPDQHRFSYQIVKRNNYKAGTNGHPYGGDYHMPRLHKFVTDAEPVEMTSYM  
p45387 DVDFGAEGNNPDQHRFTYKIVKRNNYKDD-NLHPYEDDYHNPRLHKFVTEAAPIDMTSNM  
20 orflng-1.pep DGWKYADLNKYPDRVRIGAGROYWRSDEPNRESSYHIASAYSWLVGNTFAQNGSGG  
p45387 NGSTYSDRTKYPERVRIGSGRQFWRNDQDKG-----QVAGAYHYLTAGTNHNRQAGN  
25 orflng-1.pep GTVNLGSEKIKHSPYGFPTGGSFGDSGSPFIYDAQKQKWLINGVLQTGNPYIGKSNFG  
p45387 GYSYLGGDVRKAGEYGPLIAGSKGDSGSPFIYDAEKQKWLINGILREGNPFEGKENG  
30 orflng-1.pep QLVKRDWYDEIFAGDTHSVFYEHPQNGKYFFNDNNNGAGKIDAKHKHYSLPYRLKTRTV  
p45387 QLVKRSYF-DEIFERDLHTSLYTRAGNGVYTISGNDNGQGSITQKS---GIPSEIK---I  
35 orflng-1.pep QLFNVSLSETAREPVYHAA-GGVNSYRPRLNNGENISFIDKGGELILTSNINQGAGGLY  
p45387 TLANMSLPKEKDKVHNPRYDGPNIYSPRLNGETLYFMDQKQGSILFASDINQGAGGLY  
40 orflng-1.pep FEGNFTVSPKNNETWQAGVHISDGSVTWVKVNGVANDRLSKIGKGTLLVQAKGENQGSV  
p45387 FEGNFTVSPNSNQWQAGIHVSENSTVTWVKVNGVEHDLRLSKIGKGTLLVQAKGENKGS  
45 orflng-1.pep SVGDGKVILDQQADDQGGKQAFSEIGLVSGRGTVQLNADNQFNPDKLYFGFRGGRDLNG  
p45387 SVGDGKVILEQQADDQGNKQAFSEIGLVSGRGTVQLNDDKQFDTDKFYFGFRGGRDLNG  
50 orflng-1.pep HSLSFHRIQNTDEGAMIVNHNQDKESTVTITGNKDITT-TGNN-NNLDSKKEIAYNGWFG  
p45387 HSLTFKRIQNTDEGAMIVNHNNTQAANVTITGNESIVLPNGNNINKLDYRKEIAYNGWFG  
55 orflng-1.pep EKDATKTNGRLNLNYQPEEADRTLLSGGTNLNGNITQTNGKLFFSGRPTPHAYNHLGSG  
p45387 ETDKNKHNGRLNLIYKPTTEDRTLLSGGTNLKGDITQTGKLFFSGRPTPHAYNHLNKR  
60 orflng-1.pep  
p45387  
65 orflng-1.pep  
p45387  
70 orflng-1.pep  
p45387

[illegible]



		1380	1390	1400	1410	1420	1430
	orf1ng-1.pep	AFNRYRAGIKADYSFKPAQHISITPYLSLSYTDAAAGKVRTRVNTAVLAQDFGKTRSAEW					
5	p45387	::  :   ::  :   : ::  : ::: :     :       : :					
		1300	1310	1320	1330	1340	1350
		AFNRYNAGIRVDYTFPTDNISVKPYFFVNYVDVSNANVQTTVNLTVLQQPFGRYWQKEV					
		1440	1450	1460	1469		
10	orf1ng-1.pep	GVNAEIKGFTLSLHAAAAGPQLEAQSAGIKLYRWX					
		::      :  : :      ::: :					
	p45387	GLKAEILHFQISAFISKSQGSQSLGKQNVGVKLYRW					
		1360	1370	1380	1390		

Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### 15 Example 78

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 655>:

1	..AAGGTGTGGC	AATTTGTGCA	AGA.CCGCTG	CGTGCCGTCG	TGCCTGCCGA
51	CAGTTTGTAA	CCGACCGCGC	AAAAATTGAA	CCTGTTTAAG	GCGGGTGCGG
101	CAACCATTTT	GTTTTATGAA	GATCAAAATG	TCGTCAAAGG	TTTGCAAGAG
151	CAGTTCCCTG	CTTATGCCCG	TAACTTCCCC	GTTTGGGCGG	ATCAGGCAAA
201	CGCGATGGTG	CAGTATGCCG	TTTGGACGAC	ACTTGCCGCG	GTCGGCGTAG
251	GTGCAACCT	GCAACATTAC	AATCCCTTGC	CCGATGCGGC	GATTGCCAAA
301	GCGTGAATA	TCCCCGAAAA	CTGTTTGTG	CGCGCACAAA	TGGTTATCGG
351	CGGTATTGAA	GGGCGCGCAG	GTGAAAAGAC	CTTTGAACCC	GTTGCAGAAC
401	GTTTGAAGT	GTTGCGCGCA	TAA		

This corresponds to the amino acid sequence <SEQ ID 656; ORF6>:

1	..KVVQFVEXPL	RAVVPADSFE	PTAQKLNLFK	AGAATILFYE	DQNVVKGLQE
51	QFPAYAAANFP	VWADQANAMV	QYAVWTTLAA	VGVGANLQHY	NPLPDAAIAK
101	AWNIPENWLL	RAQMVIIGIE	GAAGEKTFEP	VAERLKVFGA	*

30 Further sequence analysis revealed a further partial DNA sequence <SEQ ID 657>:

1	..CTGCGTGCCG	TCGTGCCTGC	CGACAGTTTT	GAACCGACCG	CGCAAAAATT
51	GAACCTGTTT	AAGGCGGGTG	CGGCAACCAT	TTTGTTTTAT	GAAGATCAAA
101	ATGTCGTCAA	AGGTTTGCAG	GAGCAGTTCC	CTGCTTATGC	CGCTAACTTC
151	CCCGTTTGGG	CGGATCAGGC	AAACGCGATG	GTGCAGTATG	CCGTTTGGAC
201	GACACTTGCC	GCGGTCGGCG	TAGGTGCAAA	CCTGCAACAT	TACATCCCT
251	TGCCCCGATG	GGCGATTGCC	AAAGCGTGGA	ATATCCCCGA	AAACTGGTTG
301	TTGCGCGCAC	AAATGGTTAT	CGGCGGTATT	GAAGGGGCGG	CAGGTGAAAA
351	GACCTTTGAA	CCCCTTGCAG	AACGTTTGAA	AGTGTTGCGG	GCATAA

This corresponds to the amino acid sequence <SEQ ID 658; ORF6-1>:

1	..LRAVVPADSF	EPTAQKLNLF	KAGAATILFY	EDQNVVKGLQ	EQFPAYAAANF
51	PVWADQANAM	VQYAVWTTLA	AVGVGANLQH	YNPLPDAAIA	KAWNIPENWL
101	LRAQMVIIGGI	EGAAGEKTFE	PVAERLKVFG	A*	

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

45 ORF6 shows 98.6% identity over a 140aa overlap with an ORF (ORF6a) from strain A of *N.meningitidis*:

	orf6.pep	10	20	30
		KVVQFVEXPLRAVVPADSFEPTAQKLNLFK		
50	orf6a	QIVEHAVLHTPSSFNSQSARVVVLFGEEDKVVQFVEDALRAVVPADSFEPTAQKLNLFK		

-396-

		40	50	60	70	80	90
5	orf6.pep	AGAATILFYEDQNVVKGLQE	QFPAYAA	NFPVWADQANAMVQYAVWTT	LAAVGVGANLQHY		
	orf6a	AGAATILFYEDQNVVKGLQE	QFPAYAA	NFPVWADQANAMVQYAVWTT	LAAVGVGANLQHY		
		100	110	120	130	140	150
10	orf6.pep	NPLPDAAIAKAWNIPENWLL	LRAQM	VIGGIEGAAGEKT	FEPVAERLKVFGAX		
	orf6a	NPLPDAAIAKAWNIPENWLL	LRAQM	VIGGIEGAAGEKT	FEPVAERLKVFGAX		
		160	170	180	190	200	

The complete length ORF6a nucleotide sequence <SEQ ID 659> is:

15	1	ATGACCCGTC	AATCTCTGCA	ACAGGCTGCC	GAAAGCCGCC	GTTCCATTTA
	51	TTCGTTAAAT	AAAAATCTGC	CCGTCGGCAA	AGATGAAATC	GTCCAAATCG
	101	TCGAACACGC	CGTTTTCAC	ACACCTTCTT	CGTTCAATTC	CCAATCTGCC
	151	CGTGTGGTCG	TGCTGTTTGG	CGAAGAGCAT	GATAAGGTGT	GGCAATTTGT
	201	CGAAGACGCG	CTGCGTGCCG	TCGTGCCTGC	CGACAGTTTT	GAACCGACCG
20	251	CGCAAAATTT	GAACCTGTTT	AAGGCGGGTG	CGGCAACTAT	TTTGTTTTAT
	301	GAAGATCAAA	ATGTCGTCAA	AGGTTTGCAG	GAGCAGTTCC	CTGCTTATGC
	351	CGCCAACCTT	CCCGTTTGGG	CGGACCAGGC	GAACGCGATG	GTGCAGTATG
	401	CCGTTTGGAC	GACACTTGCC	GCGGTCGGCG	TAGGTGCAAA	CCTGCAACAT
	451	TACAATCCCT	TGCCCGATGC	GCGGATTGCC	AAAGCGTGGA	ATATCCCCGA
25	501	AAACTGGTTG	TTGCGCGCAC	AAATGGTTAT	CGGCGGTATT	GAAGGGGCGG
	551	CAGGTGAAAA	GACCTTTGAA	CCAGTTGCAG	AACGTTTGAA	AGTGTTCCGG
	601	GCATAA				

This is predicted to encode a protein having amino acid sequence <SEQ ID 660>:

30	1	MTRQSLQQAA	ESRRSIYSLN	KNLPVGKDEI	VQIVEHAVLH	TPSSFNSQSA
	51	RVVVLFGEEH	DKVWQFVEDA	LRAVVPADSF	EPTAQKLNLF	KAGAATILFY
	101	EDQNVVKGLQ	EQFPAYAA	NFPVWADQANAM	VQYAVWTTLA	AVGVGANLQH
	151	YNPLPDAAIA	KAWNIPENWL	LRAQM	VIGGIEGAAGEKT	FEPVAERLKVFG
	201	A*				

ORF6a and ORF6-1 show 100.0% identity in 131 aa overlap:

		50	60	70	80	90	100
	orf6a.pep	TPSSFNSQSARVVVLFGE	EHDKVWQFVEDALRAVVP	ADSF	EPTAQKLNLF	KAGAATILFY	
	orf6-1						
40							
	orf6a.pep	EDQNVVKGLQE	QFPAYAA	NFPVWADQANAMVQYAVWTT	LAAVGVGANLQHY	NPLPDAAIA	
45	orf6-1	EDQNVVKGLQE	QFPAYAA	NFPVWADQANAMVQYAVWTT	LAAVGVGANLQHY	NPLPDAAIA	
		40	50	60	70	80	90
	orf6a.pep	KAWNIPENWLL	LRAQM	VIGGIEGAAGEKT	FEPVAERLKVFGAX		
50	orf6-1	KAWNIPENWLL	LRAQM	VIGGIEGAAGEKT	FEPVAERLKVFGAX		
		100	110	120	130		

Homology with a predicted ORF from *N.gonorrhoeae*

ORF6 shows 95.7% identity over a 140aa overlap with a predicted ORF (ORF6ng) from *N.gonorrhoeae*:

60	orf6.pep		KVWQFVEXPLRAVVPADSF	EPTAQKLNLFK	30
	orf6ng	SNVSLMSNPTVLRMGLPLYIASLR	GAIYKVWQFVEDALRAVVPADSF	EPTAQKLNLFK	64

```

orf6.pep    AGAATILFYEDQNVVKGLQEQQFPAYAAANFPVWADQANAMVQYAVWTTLAAVGVGANLQHY  90
|||||:|||||
orf6ng      AGAATILFYEDQNVVKGLQEQQFPAYAAANFPVWADQANAMVQYAVWTTLAAVGAGANLQHY 124

5  orf6.pep    NPLPDAAIAKAWNIPENWLLRAQMVIIGGIEGAAGEKTFEPVAERLKVFGA  140
|||||:|||||
orf6ng      NPLPDVAIAKAWNIPENWLLRAQMVIIGGIEGAAGEKVFEFVAERLKVFGA  174

```

The complete length ORF6ng nucleotide sequence <SEQ ID 661> was identified as:

```

10      1  ATGCCGTTG  CGTCAAATGT  CAGCTTGGAT  ATGTCCAATC  CTACGGTGTT
      51  ACGCATGGGA  TTACCCTTAT  ATATTGCGTC  CCTAAGAAGG  GGCGCAATAT
     101  ATAAGGTGTG  GCAATTTGTC  GAAGACGCGC  TGCCTGCCGT  CGTGCTTGCC
     151  GACAGTTTGT  AACCGACCGC  GCAAAAATTG  AAGCTGTTTA  AGGCGGGCGC
     201  GGCAACCATT  TTGTTTTATG  AAGATCAAAA  TGTCGTCAAA  GGTTTGCAGG
     15  251  AGCAGTTCCC  TGCTTATGCC  GCCAACTTTC  CCGTTTGGGC  GGACCAGGCG
     301  AACGCTATGG  TACAGTATGC  CGTCTGGACG  ACACTTGCCG  CGGTGCGTGC
     351  AGGTGCAAT  CTGCAACATT  ACAACCCCTT  GCCCGATGTG  GCGATTGCTA
     401  AAGCGTGGAA  TATTCGCGAA  AACTGGCTGT  TGCGCGCGCA  AATGTTTATC
     451  GGTGGTATG  AAGGGGcggc  aggtgaaaaa  gtctttgaac  CCGTTGCgga
     501  acgtttgAAA  GTGTTCCGCG  CATAA

```

This encodes a protein having amino acid sequence <SEQ ID 662>:

```

      1  MAVASNVSLD  MSNPTVLRMG  LPLYIASLRR  GAIYKVWQFV  EDALRAVVPA
     51  DSFEPTAQL  KLFKAGAATI  LFYEDQNVVK  GLQEQQFAYA  ANFPVWADQA
     101  NAMVQYAVWT  TLAAVGAGAN  LQHYNPLPDV  AIAKAWNIPE  NWLLRAQMVI
     151  GGIEGAAGEK  VFEPVAERLK  VFGA*

```

ORF6ng and ORF6-1 show 96.9% identity in 131 aa overlap:

```

      10      20      30
orf6-1.pep    LRAVVPA DSFEPTAQLNLFKAGAATILFY
      30  orf6ng      PTVLRMGLPLYIASLRRGAIYKVWQFVEDALRAVVPA DSFEPTAQLKLFKAGAATILFY
      20      30      40      50      60      70
      35  orf6-1.pep    EDQNVVKGLQEQQFPAYAAANFPVWADQANAMVQYAVWTTLAAVGVGANLQHYNPLPDAAIA
      orf6ng      EDQNVVKGLQEQQFPAYAAANFPVWADQANAMVQYAVWTTLAAVGAGANLQHYNPLPDVAIA
      80      90      100      110      120      130
      40  orf6-1.pep    KAWNIPENWLLRAQMVIIGGIEGAAGEKTFEPVAERLKVFGAX
      orf6ng      KAWNIPENWLLRAQMVIIGGIEGAAGEKVFEFVAERLKVFGAX
      140      150      160      170

```

It is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 79

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 663>

```

50      1  ..GGCTACAACT  ACCTGTTTCGC  GCGCGGCAGC  CGCATCGCCA  ACTACCAAAT
     51  CAACGGCATC  CCCGTTGCCG  ACGCGCTGGC  CGATACGGG  CAATGCCAAC
     101  ACCGCCGCT  ATGAGCGCGT  AGAAGTCGTG  CGCGGCGTGG  CGGGGCTGCT
     151  GGACGGCAGC  GGCGAGCCTT  CCGCCACCGT  CAATCTGGTG  CGCAAACGCC
     201  TGACCCGCAA  GCCATTGTTT  GAAGTCCGCG  CCGAAGCgGG  CAACCGcAAA
     55  251  CATTCGGGC  TGGACGCGGA  CGTATCGGGC  AGCCTGAACA  CCGAAG.c rC
     301  rCTGCGCgGC  CGCCTGTTTT  CCAcCTTCGG  ACGCGCGCGC  TCGTGGCGGC

```

5  
351 GGC CGCAACG CAGCCGskAT GCCGA ACTCT ACGGCATTTT GGAATACGAC  
401 ATCGCACCGC AAACCCGCGT CCACGCAGC ATGGACTACC AGCAGGCGAA  
451 AGAAACCGCC GACGCGCCGC TCAGCTACGC CGTGACGAC AGCCAAGGTT  
501 ATGCCACCGC CTTCCGCCCC AAAGACAACC CCGCCACAAA TTGGGCGAAC  
551 AGCCACCACC GTGCGCTCAA CCTGTTCCGC GGCATCGAAC ACCGCTTCAA  
601 CCAAGACTGG AAAC TCAAAG CCGAATACGA CTAC..

This corresponds to the amino acid sequence <SEQ ID 664; ORF23>:

10  
1 ..GYN YLFARGS RIANYQINGI PVADALADTG NANTAAYERV EVVRGVAGLL  
51 DGTGEP SATV NLVRKRLTRK PLFEVRAEAG NRKHFGLDAD VSGSLNTEXX  
101 LRGR LVSTFG RGDSWRRRER SRXAELYGIL EYDIAPQTRV HAXMDYQQAK  
151 ETADAPLSYA VYDSQGYATA FGPKDNPATN WANSXHRALN LFAGIEHRFN  
201 QDWKLKAEYD Y..

Further work revealed the complete nucleotide sequence <SEQ ID 665>:

15  
1 ATGACACGCT TCAAATATTC CCTGCTGTTT GCCGCCCTGT TGCCCCGTGTA  
51 CGCGCAGGCC GATGTTTCTG TTTCAGACGA CCCCCAACCG CAGGAAAGCA  
101 CTGAATTGCC GACCATCACC GTTACCGCCG ACCGCACCGC GAGTTCCAAC  
151 GACGGCTACA CTGTTTCCGG CACGCACACC CCGCTCGGGC TGCCCCATGAC  
201 CCTGCGCGAA ATCCCGCAGA GCGTCAGCGT CATCACATCG CAACAAATGC  
251 GCGACCAAAA CATCAAAACG CTCGACCGCG CCCTGTTGCA GGCGACCGGC  
20 301 ACCAGCCGCC AGATTTACGG CTCCGACCGC GCGGGCTACA ACTACCTGTT  
351 CGCGCGCGGC AGCCGCATCG CCAACTACCA AATCAACGGC ATCCCCGTTG  
401 CCGACGCGCT GGCGGATACG GGCAATGCCA ACACCGCCG CTATGAGCGC  
451 GTAGAAGTCG TGGCGGCGT GCGGGGGCTG CTGGACGGCA CGGGCGAGCC  
501 TTCCGCCACC GTCAATCTGG TGGCAAAACG CCTGACCGC AAGCCATTGT  
25 551 TTGAAGTCCG CGCCGAAGCG GGCAACCGCA AACATTTCCG GCTGGACGCG  
601 GACGTATCGG GCAGCCTGAA CACCGAAGGC ACGCTGCGCG GCCGCTGGT  
651 TTCCACCTTC GGACGCGGCG ACTCGTGGCG GCGGCGCGAA CGCAGCCGCG  
701 ATGCCGAAC CTACGGCATT TTGAATACG ACATCGCACC GCAAACCGCG  
751 GTCCACGCAG GCATGGACTA CCAGCAGCGC AAAGAAACCG CCGACGCGCC  
30 801 GCTCAGTAC GCGGTGTACG ACAGCCAAGG TTATGCCACC GCCTTCGGCC  
851 CGAAAGACAA CCGCGCCACA AATTGGGCGA ACAGCCGCCA CCGTGCGCTC  
901 AACCTGTTCC CCGGCATCGA ACACCGCTTC AACCAAGACT GGAAACTCAA  
951 AGCCGAATAC GACTACACCC GCAGCCGCTT CCGCCAGCCC TACGGCGTAG  
1001 CAGGCGTGCT TTCCATCGAC CACAACACCG CCGCCACCGA CCTGATTCCC  
35 1051 GGTATTGGC ACGCCGACCC GCGCACCAC AGCGCCAGCG TGTCATTGAT  
1101 CGGCAAATAC CGCTGTTTCG GCGCGAACA CGATTTAATC GCGGGTATCA  
1151 ACGGTTACAA ATACGCCAGC AACAAATACG GCGAACGCAG CATCATCCCC  
1201 AACGCCATTC CCAACGCCTA CGAATTTTCC CGCACGGGTG CCTACCCGCA  
1251 GCCTGCATCG TTTGCCCAA CCATCCCGCA ATACGGCACC AGGCGGCAAA  
40 1301 TCGGCGGCTA TCTCGCCACC CGTTTCCGCG CCGCCGACAA CCTTTCGCTG  
1351 ATTTTGGGCG GACGATACAC CCGTTACCGC ACCGGCAGCT ACGACAGCCG  
1401 CACACAAGGC ATGACCTATG TGTCCGCCAA CCGTTTACC CCCTACACAG  
1451 GCATCGTGTT CGACCTGACC GGCAACCTGT CTCTTTACGG CTCGTACAGC  
45 1501 AGCCTGTTTC TCCCGCAAT GCAAAAAGAC GAACACGGCA GCTACCTGAA  
1551 ACCCGTAACC GGCAACAATC TGGAGCCGG CATCAAAGGC GAATGGCTTG  
1601 AAGGCCGTCT GAACGCATCC GCCGCCGTGT ACCGCGCCCG TAAAAACAAC  
1651 CTCGCCACCG CAGCAGGACG CGACCCGAGC GGCAACACCT ACTACCGCGC  
1701 CGCCAACCAA GCCAAAACCC ACGGCTGGGA AATCGAAGTC GGCGGCCGCA  
50 1751 TCACGCCCGA ATGGCAGATA CAGGCAGGTT ACAGCCAAAG CAAAACCCGC  
1801 GACCAAGACG GCAGCCGCTT GAACCCGAC AGCGTACCCG AACGCAGCTT  
1851 CAACTCTTC ACTGCCTACC ACTTTGCCCC CGAAGCCCCC AGCGGCTGGA  
1901 CCATCGGCGC AGGGGTGCGC TGGCAGAGCG AAACCCACAC CGACCTGCC  
1951 ACGTCCGCA TCCCAACCC CGCCGCCAAA GCCCGCGCCG CCGACAACAG  
55 2001 CCGCCAAAA GCCTACGCCG TCGCCGACAT CATGGCGCGT TACCGCTTCA  
2051 ATCCGCGCGC CGAACTGTGC CTGAACGTGG ACAATCTGTT CAACAAACAC  
2101 TACCGCACCC AGCCCGACCG CCACAGCTAC GGCGCACTGC GGACAGTGAA  
2151 CGCGGCGTTT ACCTATCGGT TTAAATAA

This corresponds to the amino acid sequence <SEQ ID 666; ORF23-1>:

60  
1 MTRFKYSLLF AALLPVYAQA DVSVD DPKP QESTELPTIT VTADRTASSN  
51 DGYTVSGTHT PLGLPMTLRE IPQSVSVITS QQMRDQNIKT LDRALLQATG  
101 TSRQIYGS DR AGYN YLFARG SRIANYQING IPVADALADT GNANTAAYER  
151 VEVVRGVAGL LDGTGEP SAT VNLVRKRLTR KPLFEVRAEA GNRKHFGLDA  
201 DVSGSLNTEG TLRGR LVSTF GRGDSWRRRE RSRDAELYGI LEYDIAPQTR

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65  
70

```

251 VHAGMDYQQA KETADAPLSY AVYDSQGYAT AFGPKDNPAT NWANSRHRAL
301 NLFAGIEHRF NQDWKLKAEY DYTRSFRQPY YGVAGVLSID HNTAATDLIP
351 GYWHADPRTH SASVSLIGKY RLFGREHDLI AGINGYKYAS NKYGERSIIP
401 NAIPNAYEFS RTGAYPQPAS FAQTIPQYGT RRQIGGYLAT RFRAADNLSL
451 ILGGRYTRYR TGSYDSRTQG MTYVVSANFT PYTGIVFDLT GNLSLYGSYS
501 SLFVPQSQKD EHGSYLKPVT GNNLEAGIKG EWLEGRLNAS AAVYRARKNN
551 LATAAGRDPS GNTYYRAANQ AKTHGWEIEV GGRITPEWQI QAGYSQSKTR
601 DQDGSRLNPD SVPERSFKLF TAYHFAPEAP SGWTIGAGVR WQSEHTDPA
651 TLRIPNPAK ARAADNSRQK AYAVADIMAR YRFNPRAELS LNVDNLFNKH
701 YRTQPDHRSY GALRTVNAAF TYRFK*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with the ferric-pseudobactin receptor PupB of *Pseudomonas putida* (accession number P38047)

ORF23 and PupB protein show 32% aa identity in 205aa overlap:

15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65  
70  
75  
80  
85  
90  
95  
100  
105  
110  
115  
120  
125  
130  
135  
140  
145  
150  
155  
160  
165  
170  
175  
180  
185  
190  
195  
200  
205  
210  
215  
220  
225  
230  
235  
240  
245  
250  
255  
260  
265  
270  
275  
280  
285  
290  
295  
300  
305  
310  
315  
320  
325  
330  
335  
340  
345  
350  
355  
360  
365  
370  
375  
380  
385  
390  
395  
400  
405  
410  
415  
420  
425  
430  
435  
440  
445  
450  
455  
460  
465  
470  
475  
480  
485  
490  
495  
500  
505  
510  
515  
520  
525  
530  
535  
540  
545  
550  
555  
560  
565  
570  
575  
580  
585  
590  
595  
600  
605  
610  
615  
620  
625  
630  
635  
640  
645  
650  
655  
660  
665  
670  
675  
680  
685  
690  
695  
700  
705  
710  
715  
720  
725  
730  
735  
740  
745  
750  
755  
760  
765  
770  
775  
780  
785  
790  
795  
800  
805  
810  
815  
820  
825  
830  
835  
840  
845  
850  
855  
860  
865  
870  
875  
880  
885  
890  
895  
900  
905  
910  
915  
920  
925  
930  
935  
940  
945  
950  
955  
960  
965  
970  
975  
980  
985  
990  
995  
1000

```

Orf23 6 FARGSRIANYQINGIPVADALADTGNANTAAYERVEVVRGVAGLLDGTGEP SATVNLRK 65
++RG I NY+++G+P + L D + + A ++RVE+VRG GL+ G G PSAT+NL+RK
PupB 215 WSRGFAIQNYEVDGVPTSTRL-DNYSQSMAMFDRVEIVRGATGLISGMGNPSATINLIRK 273

Orf23 66 RLTRKPLFEVRAEAGNRKHFGLDADVSGSLNTEXXLRGLRVSTFXXXXXXXXXXXXXAE 125
R T + + EAGN +G DVSG L +RGR V+ +
PupB 274 RPTAEAQASITGEAGNWDRYGTGFVSGPLTETGNIRGRFVADYKTEKAWIDRYNQSQL 333

Orf23 126 LYGILEYDIAPQTRVHAXMDYQQAKETADAPLSYAVYD--SQGYATAFGPKDNPATNWAN 183
+YGI E+D++ T + Y + D+PL + S G T N A +W+
PupB 334 MYGITEFDLSEDTLLTVGFSY--LRSDIDSPLRSGLPTRFSTGERTNLKRS LNAAPDWSY 391

Orf23 184 SHHRALNLFAGIEHRFNQDWKLKAE 208
+ H + F IE + W K E
PupB 392 NDHEQTSFFTSIEQQLGNGWSGKIE 416

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF23 shows 95.7% identity over a 211aa overlap with an ORF (ORF23a) from strain A of *N. meningitidis*:

35  
40  
45  
50  
55  
60  
65  
70  
75  
80  
85  
90  
95  
100  
105  
110  
115  
120  
125  
130  
135  
140  
145  
150  
155  
160  
165  
170  
175  
180  
185  
190  
195  
200  
205  
210  
215  
220  
225  
230  
235  
240  
245  
250  
255  
260  
265  
270  
275  
280  
285  
290  
295  
300  
305  
310  
315  
320  
325  
330  
335  
340  
345  
350  
355  
360  
365  
370  
375  
380  
385  
390  
395  
400  
405  
410  
415  
420  
425  
430  
435  
440  
445  
450  
455  
460  
465  
470  
475  
480  
485  
490  
495  
500  
505  
510  
515  
520  
525  
530  
535  
540  
545  
550  
555  
560  
565  
570  
575  
580  
585  
590  
595  
600  
605  
610  
615  
620  
625  
630  
635  
640  
645  
650  
655  
660  
665  
670  
675  
680  
685  
690  
695  
700  
705  
710  
715  
720  
725  
730  
735  
740  
745  
750  
755  
760  
765  
770  
775  
780  
785  
790  
795  
800  
805  
810  
815  
820  
825  
830  
835  
840  
845  
850  
855  
860  
865  
870  
875  
880  
885  
890  
895  
900  
905  
910  
915  
920  
925  
930  
935  
940  
945  
950  
955  
960  
965  
970  
975  
980  
985  
990  
995  
1000

```

orf23.pep 10 20 30
GYNLYFARGSRIANYQINGIPVADALADTG
|||||
orf23a QMRDQNIKALDRALLQATGTSRQIYGSDRAGYNLYFARGSRIANYQINGIPVADALADTG
90 100 110 120 130 140

orf23.pep 40 50 60 70 80 90
NANTAAYERVEVVRGVAGLLDGTGEP SATVNLRKRLTRKPLFEVRAEAGNRKHFGLDAD
|||||
orf23a NANTAAYERVEVVRGVAGLLDGTGEP SATVNLRKRPTRKPLFEVRAEAGNRKHFGLGAD
150 160 170 180 190 200

orf23.pep 100 110 120 130 140 150
VSGSLNTEXXLRGLRVSTFGRGDSWRRRERSRAXEALYGILEYDIAPQTRVHAXMDYQQAK
|||||:| :|||||:|||||:|||||
orf23a VSGSLNAEGTLRGLRVSTFGRGDSWRQRERSRDAELYGILEYDIAPQTRVHAGMDYQQAK
210 220 230 240 250 260

orf23.pep 160 170 180 190 200 210
ETADAPLSYAVYDSQGYATAFGPKDNPATNWANSHHRALNLFAGIEHRFNQDWKLKAEYD
|||||:|||||:|||||:|||||:|||||
orf23a ETADAPLSYAVYDSQGYATAFGPKDNPATNWANSHHRALNLFAGIEHRFNQDWKLKAEYD
270 280 290 300 310 320

orf23.pep Y
|
orf23a YTRSFRQPYGVAGVLSIDHNTAATDLIPGYWHADPRTHSASVSLIGKYRLF GREHDLIA
330 340 350 360 370 380

```

The complete length ORF23a nucleotide sequence <SEQ ID 667> is:

5  
10  
15  
20  
25  
30  
35  
40  
45

```

1  ATGACACGCT TCAAATATTC CCTGCTGTTT GCCGCCTGT TGCCCGTGTA
51  CGCGCAGGCC GATGTTTCTG TTTCAGACGA CCCAAAACCG CAGGAAAGCA
101 CTGAATTGCC GACCATCACC GTTACCGCCG ACCGCACCGC GAGTTCCAAC
151 GACGGCTACA CTGTTTCCGG CACGCACACC CCGCTCGGGC TGCCCATGAC
201 CCTGCGCGAA ATCCCGCAGA GCGTCAGCGT CATCACATCG CAACAAATGC
251 GCGACCAAAA CATCAAAGCG CTCGACCGCG CCCTGTTGCA GGCGACCGGC
301 ACCAGCCGCG AGATTTACGG CTCCGACCGC GCGGGCTACA ACTACCTGTT
351 CGCGCGCGGC AGCCGCATCG CCAACTACCA AATCAACGGC ATCCCCGTTG
401 CCGACGCGCT GGCGGATACG GGCAATGCCA ACACCGCCGC CTATGAGCGC
451 GTAGAAGTCG TCGCGGCGGT GCGGGGCTG CTGGACGGCA CCGGGGAGCC
501 TTCCGCCACC GTCAATCTGG TCGCAAACG CCCGACCCGC AAGCCATTGT
551 TTGAAGTCGG CGCCGAAGCG GGCAACCGCA AACATTTGCG GCTGGGCGCG
601 GACGTATCGG GCAGCCTGAA TGCCGAAGGC ACGCTGCGCG GCCGCTGTTG
15  651 TTCCACCTTC GGACGCGGCG ACTCGTGGCG GCAGCGCGAA CGCAGCCGCG
701 ATGCCGAACT CTACGGCATT TTGAATACG ACATCGCACC GCAAAACCGC
751 GTCCACGCAG GCATGGACTA CCAGCAGGCG AAAGAAACCG CCGACGCGCC
801 GCTCAGCTAC GCCGTGTACG ACAGCCAAGG TTATGCCACC GCCTTCGGCC
851 CGAAAGACAA CCCC GCCACA AATTGGGCGA ACAGCCGCCA CCGTGCCTC
20  901 AACCTGTTTC CCGGCATCGA ACACCGCTTC AACCAAGACT GGAAACTCAA
951 AGCCGAATAC GACTACACCC GCAGCCGCTT CCGCCAGCCC TACGGCGTAG
1001 CAGGCGTGCT TTCCATCGAC CACAACACCG CCGCCACCGA CCTGATTCCT
1051 GGTATTGCG ACGCCGACCC GCGCACCCAG AGCGCCACGG TGTCATTAAAT
1101 CGGCAAATAC CGCTGTTCG GCGCGAACA CGATTTAATC GCGGGTATCA
25  1151 ACGGTACAA ATACGCCAGC AACAAATACG GCGAACGCGA CATCATCCCC
1201 AACGCCATT CCAACGCCCTA CGAATTTTCC CGCACGGGTG CCTACCCGCA
1251 GCCTGCATCG TTTGCCCAA CCATCCCGCA ATACGGCACC AGGCGGCAAA
1301 TCGCGGGCTA TCTCGCCACC CGTTTCCGCG CCGCCGACAA CCTTTCGCTG
1351 ATACTCGGCG GCAGATACAG CCGTTACCGC ACCGGCAGCT ACGACAGCCG
30  1401 CACACAAGGC ATGACCTATG TGTCCGCCAA CCGTTTCACC CCTACACAG
1451 GCATCGTGTT CGACCTGACC GGCAACCTGT CGCTTTACGG CTCGTACAGC
1501 AGCCTGTTTC TCCCGCAATC GCAAAAAGAC GAACACGGCA GCTACCTGAA
1551 ACCCGTAACC GGCAACAATC TGAAGCCGG CATCAAAGGC GAATGGCTTG
1601 AAGGCCGTCT GAACGCATCC GCCGCCGTGT ACCGCGCCCG TAAAAACAAC
35  1651 CTCGCCACCG CAGCAGGACG CGACCCGAGC GGCAACACCT ACTACCGCGC
1701 CGCCAACCAA GCCAAAACCC ACGGCTGGGA AATCGAAGTC GCGCGCCGCA
1751 TCACGCCCGA ATGGCAGATA CAGGCAGGTT ACAGCCAAAG CAAAACCCGC
1801 GACCAAGACG GCAGCCGCCT GAACCCCGAC AGCGTACCGG AACGCAGCTT
1851 CAAACTCTTC ACTGCCTACC ACTTTGCCCC CGAAGCCCCC AGCGGTGGA
40  1901 CCATCGGCGC AGGCGTGCGC TGGCAGAGCG AAACCCACAC CGACCTGGC
1951 ACGTCCGCA TCCCAACCC CGCCGCCAAA GCCCGCGCCG CCGACAACAG
2001 CCGCCAAAAA GCCTACGCGG TCGCCGACAT CATGGCGCGT TACCGCTTCA
2051 ATCCGCGCGC CGAACTGTCT CTGAACGTGG ACAATCTGTT CAACAAACAC
2101 TACCGCACCC AGCCCGACCG CCACAGCTAC GCGCAGTGC GGACAGTGAA
45  2151 CGCGCGGTTT ACCTATCGGT TTAATAAA

```

This encodes a protein having amino acid sequence <SEQ ID 668>:

50  
55  
60

```

1  MTRFKYSLLF AALLPVYAQA DVSVDDBPKP QESTELPTIT VTADRTASSN
51  DGYTVSGTHT PLGLPMTLRE IPQSVSVITS QQMRDQNIKA LDRALLQATG
101 TSRQIYGS DR AGYNILFARG SRIANYQING IPVADALADT GNANTAAYER
151 VEVVRGVAGL LDGTGEP SAT VNLVRKR PTR KPLFEVRAEA GNRKHFLGA
201 DVSGSLNAEG TLRGRLVSTF GRGDSWRQRE RSRDAELYGI LEYDIAPQTR
251 VHAGMDYQQA KETADAPLSY AVYDSQGYAT AFGPKDNPAT NWANSRHRAL
301 NLFAGIEHRF NQDWKLKAEY DYTRSRFRQP YGVAGVLSID HNTAATDLIP
351 GYWHADPRTH SASVSLIGKY RLFGREHDLI AGINGYKYAS NKYGERSIIP
401 NAIPNAYEFS RTGAYPQPAS FAQTIPQYGT RRQIGGYLAT RFRAADNLSL
451 ILGGYRSRYR TGSYDSRTQG MTYVSANRET PYTGIVFDLT GNLSLYGSYS
501 SLFVPQSQKD EHGSYLK PVT GNNLEAGIKG EWLEGRNAS AAVYRARKNN
551 LATAAGRDP S GNTYYRAANQ AKTHGWEIEV GGRITPEWQI QAGYSQSKTR
601 DDGGSRLNPD SVPERSFKLF TAYHFAPEAP SGWTIGAGVR WQSETHDPA
651 TLRIPNPAK ARAADNSRQK AYAVADIMAR YRFPRAELS LNVNLFNKH
701 YRTQPRHSY GALRTVNAAF TYRFK*

```

ORF23a and ORF23-1 show 99.2% identity in 725 aa overlap:

orf23a.pep      10      20      30      40      50      60  
MTRFKYSLLF AALLPVYAQA DVSVDDBPKP QESTELPTIT VTADRTASSN DGYTVSGTHT

	orf23-1	 MTRFKYSLLF AALLPVYAQADVSVSDPKPQESTELPTITVTADRTASNDGYTVSGTHT 10 20 30 40 50 60
5	orf23a.pep	70 80 90 100 110 120 PLGLPMTLREIPQSVSVITSQQMRDQNIKALDRALLQATGTSRQIYGSDRAGYNLYFARG      :
10	orf23-1	70 80 90 100 110 120 PLGLPMTLREIPQSVSVITSQQMRDQNIKTLDRALLQATGTSRQIYGSDRAGYNLYFARG
	orf23a.pep	130 140 150 160 170 180 SRIANYQINGIPVADALADTGNANTAAYERVEVVRGVAGLLDGTGEP SATVNLRKRPRTR 
15	orf23-1	130 140 150 160 170 180 SRIANYQINGIPVADALADTGNANTAAYERVEVVRGVAGLLDGTGEP SATVNLRKRRLTR
	orf23a.pep	190 200 210 220 230 240 KPLFEVRAEAGNRKHFGLGADVSGSLNAEGTLRGRLVSTFGRGDSWRQRERSRDAELYGI 
20	orf23-1	190 200 210 220 230 240 KPLFEVRAEAGNRKHFGLGADVSGSLNTEGTLRGRLVSTFGRGDSWRRRERSRDAELYGI
	orf23a.pep	250 260 270 280 290 300 LEYDIAPOQTRVHAGMDYQQAKETADAPLSYAVYDSQGYATAFGPKDNPATNWANSRHRAL 
25	orf23-1	250 260 270 280 290 300 LEYDIAPOQTRVHAGMDYQQAKETADAPLSYAVYDSQGYATAFGPKDNPATNWANSRHRAL
	orf23a.pep	310 320 330 340 350 360 NLFAGIEHREFNQDWKLKAEYDYTRSFRQPYGVAGVLSIDHNATAATDLIPGYWHADPRTH 
30	orf23-1	310 320 330 340 350 360 NLFAGIEHREFNQDWKLKAEYDYTRSFRQPYGVAGVLSIDHNATAATDLIPGYWHADPRTH
	orf23a.pep	370 380 390 400 410 420 SASVSLIGKYRLFGREHDLIAGINGYKYASNKYGERSIIPNAIPNAYEF SRTGAYPQPAS 
35	orf23-1	370 380 390 400 410 420 SASVSLIGKYRLFGREHDLIAGINGYKYASNKYGERSIIPNAIPNAYEF SRTGAYPQPAS
	orf23a.pep	430 440 450 460 470 480 FAQTIPQYGTTRRQIGGYLATRFRAADNLSLILGGRYTRYRTGSYDSRTQGMTYVSANRFT 
40	orf23-1	430 440 450 460 470 480 FAQTIPQYGTTRRQIGGYLATRFRAADNLSLILGGRYTRYRTGSYDSRTQGMTYVSANRFT
	orf23a.pep	490 500 510 520 530 540 PYTGIVFDLTGNLSLYGSYSSLFVPQSQKDEHGSYLPVTKGNNLEAGIKGEWLEGRNLAS 
45	orf23-1	490 500 510 520 530 540 PYTGIVFDLTGNLSLYGSYSSLFVPQSQKDEHGSYLPVTKGNNLEAGIKGEWLEGRNLAS
	orf23a.pep	550 560 570 580 590 600 AAVYRARKNNLATAAGRDPSGNTYYRAANQAKTHGWEIEVGGRITPEWQIQAGYSQSKTR 
50	orf23-1	550 560 570 580 590 600 AAVYRARKNNLATAAGRDPSGNTYYRAANQAKTHGWEIEVGGRITPEWQIQAGYSQSKTR
	orf23a.pep	610 620 630 640 650 660 DQDGSRLNPDSVPERSFKLFTAYHFAPEAPSGWTIGAGVRWQSETHDTPATLRIPNPAAK 
55	orf23-1	610 620 630 640 650 660 DQDGSRLNPDSVPERSFKLFTAYHFAPEAPSGWTIGAGVRWQSETHDTPATLRIPNPAAK
	orf23a.pep	670 680 690 700 710 720 ARAADNSRQKAYAVADIMARYRFNPRAELSLNVDNLFNKHYRTQPDHRSYGALRTVNAAF 
60	orf23-1	670 680 690 700 710 720 ARAADNSRQKAYAVADIMARYRFNPRAELSLNVDNLFNKHYRTQPDHRSYGALRTVNAAF
	orf23a.pep	670 680 690 700 710 720 ARAADNSRQKAYAVADIMARYRFNPRAELSLNVDNLFNKHYRTQPDHRSYGALRTVNAAF
65	orf23-1	670 680 690 700 710 720 ARAADNSRQKAYAVADIMARYRFNPRAELSLNVDNLFNKHYRTQPDHRSYGALRTVNAAF
70	orf23-1	670 680 690 700 710 720 ARAADNSRQKAYAVADIMARYRFNPRAELSLNVDNLFNKHYRTQPDHRSYGALRTVNAAF

orf23a.pep TYRFXK  
 |||||  
 orf23-1 TYRFXK

5

Homology with a predicted ORF from *N.gonorrhoeae*

ORF23 shows 93.4% identity over a 211aa overlap with a predicted ORF (ORF23.ng) from *N. gonorrhoeae*:

10	orf23.pep	GYNYLFARGSR IANYQINGIPVADALADTGNANTAAAYERVEVVRGVAGLLD	51
	orf23ng	SAVDACRIPG YNYLFARGSR IANYQINGIPVADALADTGNANTAAAYERVEVVRGVAGLPD	60
	orf23.pep	GTGEP SATVN LVRKRLTRKPLFEVRAEAGNRKHFGLDADVSGSLNTEXXLRGRLVSTFGR	111
15	orf23ng	GTGEP SATVN LVRKHPTKPLFEVRAEAGNRKHFGLDADVSGSLNAEGTLRGRLVSTFGR	120
	orf23.pep	GDSWRRRERSR XAELYGILEYDIAPQTRVHAXMDYQQAKETADAPLSYAVYDSQGYATAF	171
	orf23ng	GDSWRLQLERSR DAELYGILEYDIAPQTRVHAGMDYQQAKETADAPLSYAVYDSQGYATAF	180
20	orf23.pep	GPKDNPATNWNASHHRLNLFAGIEHRFNQDWKLKAEYDY	211
	orf23ng	GPKDNPATNWSNRNRLNLFAGIEHRFNQDWKLKAEYDYTRSRFRQPYGVAGVLSIDHS	240

The ORF23ng nucleotide sequence <SEQ ID 669> is predicted to encode a protein comprising amino acid sequence <SEQ ID 670>:

	1	SAVDACRIPG	YNYLFARGSR	IANYQINGIP	VADALADTGN	ANTAAYERVE
	51	VVRGVAGLPD	GTGEP SATVN	LVRKHPTKRP	LFEVRAEAGN	RKHFGLGADV
	101	SGSLNAEGL	RGRVSTFGR	GDSWRLQLERS	RDAELYGILE	YDIAPQTRVH
	151	AGMDYQQAKE	TADAPLSYAV	YDSQGYATAF	GPKDNPATNW	SNSNRNALNL
30	201	FAGIEHRFNQ	DWKLKAEYDY	TRSRFRQPYG	VAGVLSIDHS	TAATDLIPGY
	251	WHADPRTHSA	SMSLTGKYRL	FGREHDLIAG	INGYKYASNK	YGERSIIPNA
	301	IPNAYEFSRT	GAYPQPSSFA	QTIPOYDTRR	QIGGYLATRF	RAADNLSLIL
	351	GGRYSRYRAG	SYNSRTQGMT	YVSANRFTPY	TGIVFDLTGN	LSLYGSYSSL
35	401	FVPQLQKDEH	GSYLKPVTGN	NLEADIKGEW	LEGRNLNASAA	VYRARKNNLA
	451	TAAGRDSQGN	TYYRAANQAK	THGWEIEVGG	RITPEWQIQ	QAGYSQSKPRDQ
	501	DGSRLNPDSV	PERSEKLFET	YHLAPEAPSG	RTIGAGVRRQ	GETHTDPAAL
	551	RIPNPAAKAR	AVANSRQKAY	AVADIMARYR	FNPRTLSLN	VDNLFNKHRY
	601	TQPDHRSYGA	LRTVNAAFY	RFK*		

Further work revealed the complete nucleotide sequence <SEQ ID 671>:

40	1	ATGACACGCT	TCAAATACTC	CCTGCTTTTT	GCCGCCCTGC	TACCCGTGTA
	51	CGCGCAGGCC	GATGTTTCTG	TTTCAGACGA	CCCCAAACCG	CAGGAAAGCA
	101	CCGAATTGCC	GACCATCACC	GTTACCGCCG	ACCGCACCGC	GAGTTCCAAC
	151	GACGGCTACA	CCGTTTCCGG	CACGCACACC	CCGTTCCGGC	TGCCCATGAC
45	201	CCTGCGCGAA	ATCCCGCAGA	GCGTCAGCGT	CATCACATCG	CAACAAATGC
	251	GCGACCAAAA	CATCAAAACG	CTCGACCGCG	CCCTGTTGCA	GGCGACCGGC
	301	ACCAGCCGCC	AGATTTACGG	CTCCGACCGC	GCGGGCTACA	ACTACCTGTT
	351	CGCGCGCGGC	AGCCGCATCG	CCAACTACCA	AATCAACGGC	ATCCCCGTTG
	401	CCGACGCGCT	GGCCGATACG	GGCAATGCCA	ACACCGCGCG	CTATGAGCGC
50	451	GTAGAAGTCG	TGCGCGGCGT	GGCGGGGCTG	CCGGACGGCA	CGGGCGAGCC
	501	TTCTGCCACC	GTCAATCTGG	TACGCAACAA	CCCGACCCGC	AAGCCATTGT
	551	TTGAAGTCCG	CGCCGAAGCC	GGCAACCGCA	AACATTTCCG	GCTGGGCGCG
	601	GACGTATCGG	GCAGCCTGAA	CGCCGAAGGC	ACGCTGCGCG	GCCGCTGGGT
	651	TTCCACCTTC	GGACGCGGCG	ACTCGTGGCG	GCAGCTCGAA	CGCAGCCGCG
55	701	ATGCCGAACT	CTACGGCATT	TTGGAATACG	ACATCGCACC	GCAAAACCGC
	751	GTCCACGCAG	GCATGGACTA	CCAGCAGGCG	AAAGAAACCG	CAGACGCGCC
	801	GCTCAGCTAC	GCCGTGTACG	ACAGCCAAGG	TTATGCCACC	GCCTTCGGCC
	851	CAAAAGACAA	CCCCGCCACA	AATTGGTCGA	ACAGCCGCAA	CCGTGCGCTC
	901	AACCTGTTTC	CCGGCATAGA	ACACCGCTTC	AACCAAGACT	GGAAACTCAA
60	951	AGCCGAATAC	GACTACACCC	GTAGCCGCTT	CCGCCAGCCC	TACGGTGTGG
	1001	CAGGCGTACT	TTCCATCGAC	CACAGCACTG	CCGCCACCGA	CCTGATTCCC



5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65

```

1051 GGTATTGGC ACGCegatcc GCGCACCCAC AGCGCCAGCA TGTCATTGAC
1101 CGGCAAATAC CgcctGTTCG GCCGCGAGCA CGATTTAATC GCGGGTATCA
1151 ACGGCTACAA ATACGCCAGC AACAAATACG GCGAACGCAG CATCATTTCC
1201 AACGCCATTC CCAACGCCTA CGAATTTTCC CGCACGGGCG CCTATCCGCA
1251 GCCATCATCG TTTGCCCCAA CCATCCCGCA ATACGACACC AGGCGGCAAA
1301 TCGGCGGCTA TCTCGCCACC CGTTTCCGCG CCGCCGACAA CCTTTCGCTG
1351 ATACTCGGCG GCAGATACAG CCGCTACCGC GCAGGCAGCT ACAACAGCCG
1401 CACACAAGGC ATGACCTATG TGTCCGCCAA CCGTTTACC CCCTACACAG
1451 GCATCGTGTG CGATCTGACC GGCAACCTGT CGCTTTACGG CTCTACAGC
1501 AGCCTGTTTCG TCCCAGCAAT GCAAAAAGAC GAACACGGCA GCTACCTGAA
1551 ACCCGTAACC GGCAACAATC TGAAGCCGA CATCAAAGGC GAATGGCTTG
1601 AAGGGCGTCT GAACGCATCC GCCGCCGTGT ACCGCGCCCG TAAAAACAAC
1651 CTCGCCACCG CAGCAGGACG CGACCAGAGC GGCAACACCT ACTATCGCGC
1701 CGCCAACCAA GCCAAAACCC ACGGCTGGGA AATCGAAGTC GGCGGCCGCA
1751 TCACGCCCGA ATGGCAGATA CAGGCAGGCT ACAGCCAAAG CAAACCCCGC
1801 GACCAAGACG GCAGCCGCCT GAACCCGAC AGCGTACCCG AACGCAGCTT
1851 CAAACTCTTC ACCGCCTACC ACTTAGCCCC CGAAGCCCC AGCGGCCGGA
1901 CCATcggTGC GGGTGTGCGC CGGCAGGGCG AAACCCACAC CGACCCAGCC
1951 CGGCTCCGCA TCCCAACCC CGCCGCCAAA GCCCGCGCCG TCGCCAACAG
2001 CCGCCAGAAA GCCTACGCCG TCGCCGACAT CATGGCGCGT TACCGCTTCA
2051 ATCCGCGCAC CGAACTGTCG CTGAACGTGG ACAACCTGTT CAACAAACAC
2101 TACCGCACCC AGCCCGACCG CCACAGCTAC GGCGCACTGC GGACAGTGAA
2151 CGCGGCGTTT ACCTATCGGT TTAATAAA

```

This corresponds to the amino acid sequence <SEQ ID 672; ORF23ng-1>:

25  
30  
35  
40  
45  
50  
55  
60  
65

```

1 MTRFKYSLLE AALLPVYAQA DVSVSDDPKP QESTELPTIT VTADRTASSN
51 DGYTVSGTHT PFGLPMTLRE IPQSVSVITS QQMRDQNIKT LDRALLQATG
101 TSRIYGS DR AGYNILFARG SRIANYQING IPVADALADT GNANTAYER
151 VEVVRGVAGL PDGTGEPSAT VNLVRKHPT R KPLFEVRAEA GNRKHFLGA
201 DVSGSLNAEG TLRGRLVSTF GRGDSWRQLE RSRDAELYGI LEYDIAPQTR
251 VHAGMDYQQA KETADAPLSY AVYDSQGYAT AFGPKDNPAT NWSNSRNLAL
301 NLFAGIEHRF NQDWKLKAEY DYTRSRFRQP YGVAGVLSID HSTAATDLIP
351 GYWHADPRTH SASMSLTGKY RLFGREHDLI AGINGYKYAS NKYGRSIIIP
401 NAIPNAYEFS RTGAYPPQSS FAQTIPQYDT RRQIGGYLAT RFRAADNLSL
451 ILGGYRSRYR AGSYNSRTQG MTYVSANRET PYTGIVFDLT GNLSLYGSYS
501 SLFVPQLQKD EHGSYLKPV T GNNLEADIKG EWLEGRNLAS AAVYRARKNN
551 LATAAGRDQS GNTYRAANQ AKTHGWEIEV GGRITPEWQI QAGYSQSKPR
601 DQDGSRLNPD SVPERSFKLF TAYHLAPEAP SGRITIGAVR RQGETHTDPA
651 ALRIPNPAK ARAVANSRQK AYAVADIMAR YRFNPRTELS LNVNLFENKH
701 YRTQPDHRSY GALRTVNAAF TYRFK*

```

ORF23ng-1 and ORF23-1 show 95.9% identity in 725 aa overlap:

40  
45  
50  
55  
60  
65

```

              10      20      30      40      50      60
orf23-1.pep  MTRFKYSLLEAALLPVYAQADVSVSDDPKPQESTELPTITVTADRTASSNDGYTVSGTHT
              |||
orf23ng-1    MTRFKYSLLEAALLPVYAQADVSVSDDPKPQESTELPTITVTADRTASSNDGYTVSGTHT
              10      20      30      40      50      60

              70      80      90      100     110     120
orf23-1.pep  PLGLPMTLREIPQSVSVITSQQMRDQNIKTLDRLALQATGTSRQIYGS DRAGYNILFARG
              |:|
orf23ng-1    PFGLPMTLREIPQSVSVITSQQMRDQNIKTLDRLALQATGTSRQIYGS DRAGYNILFARG
              70      80      90      100     110     120

              130     140     150     160     170     180
orf23-1.pep  SRIANYQINGIPVADALADTGNANTAYERVEVVRGVAGLLDGTGEPSATVNLVRKRLTR
              |||
orf23ng-1    SRIANYQINGIPVADALADTGNANTAYERVEVVRGVAGLPDGTGEPSATVNLVRKHPT R
              130     140     150     160     170     180

              190     200     210     220     230     240
orf23-1.pep  KPLFEVRAEAGNRKHFLGLDADVSGSLNTEGTLRGRIVSTFGRGDSWRRRERSRDAELYGI
              |||
orf23ng-1    KPLFEVRAEAGNRKHFLGLDADVSGSLNAEGTLRGRIVSTFGRGDSWRQLERSRDAELYGI
              190     200     210     220     230     240

              250     260     270     280     290     300
orf23-1.pep  LEYDIAPQTRVHAGMDYQQA KETADAPLSYAVYDSQGYATAFGPKDNPATNWSNRHRAL

```

	orf23ng-1	LEYDIA PQTRVHAGMDYQQA KETADAPLSYAVYDSQGYATAFGPKDNPATNWSNSRNRAL	250	260	270	280	290	300
5	orf23-1.pep	NLFAGIEHRRFNQDWKLKAEYDYTRSFRQPYGVAGVLSIDHNTAATDLIPGYWHADPRTH	310	320	330	340	350	360
	orf23ng-1	NLFAGIEHRRFNQDWKLKAEYDYTRSFRQPYGVAGVLSIDHNTAATDLIPGYWHADPRTH	310	320	330	340	350	360
10	orf23-1.pep	SASVSLIGKYRLF GREHDLIAGINGYKYASNKYGERSIIPNAIPNAYEFSRTGAYPQPAS	370	380	390	400	410	420
	orf23ng-1	SASMSLTGKYRLF GREHDLIAGINGYKYASNKYGERSIIPNAIPNAYEFSRTGAYPQPSS	370	380	390	400	410	420
20	orf23-1.pep	FAQTIPQYGTFRRIQIGGYLATRFRAADNLSLILGGRYTRYRTGSYDSRTQGMYTYSANRFT	430	440	450	460	470	480
	orf23ng-1	FAQTIPQYDTRRIQIGGYLATRFRAADNLSLILGGYSRYRAGSYNSRTQGMYTYSANRFT	430	440	450	460	470	480
25	orf23-1.pep	PYTGIVFDLTGNLSLYGSYSSLFVPQSQKDEHGSYLKPVTGNNLEAGIKGEWLEGRNLAS	490	500	510	520	530	540
	orf23ng-1	PYTGIVFDLTGNLSLYGSYSSLFVPQLQKDEHGSYLKPVTGNNLEADIKGEWLEGRNLAS	490	500	510	520	530	540
30	orf23-1.pep	AAVYRARKNNLATAAGRDPSGNTYYRAANAKTHGWEIEVGGRITPEWQIQAGYSQSKTR	550	560	570	580	590	600
	orf23ng-1	AAVYRARKNNLATAAGRDQSGNTYYRAANAKTHGWEIEVGGRITPEWQIQAGYSQSKPR	550	560	570	580	590	600
35	orf23-1.pep	DQDGSRLNPDSVPERSEFKLFTAYHFAPEAPSGWTIGAGVRRQSETHTPAALRIPNPAAK	610	620	630	640	650	660
	orf23ng-1	DQDGSRLNPDSVPERSEFKLFTAYHLAPEAPSGRTIGAGVRRQGETHTDPAALRIPNPAAK	610	620	630	640	650	660
40	orf23-1.pep	ARAADNSRQKAYAVADIMARYRFNPRAELSLNVDNLFNKHYRTQPDHRSYGALRTVNAAF	670	680	690	700	710	720
	orf23ng-1	ARAVANSRQKAYAVADIMARYRFNPRTLSLNVDNLFNKHYRTQPDHRSYGALRTVNAAF	670	680	690	700	710	720
50	orf23-1.pep	TYRFXK						
	orf23ng-1	TYRFXK						

In addition, ORF23ng-1 shows significant homology with an OMP from *E.coli*:

sp|P16869|FHUE\_ECOLI OUTER-MEMBRANE RECEPTOR FOR FE(III)-COPROGEN, FE(III)-  
FERRIOXAMINE B AND FE(III)-RHODOTRULIC ACID PRECURSOR >gi|1651542|gnl|PID|d1015403  
55 (D90745) Outer membrane protein FhuE precursor [Escherichia coli  
>gi|1651545|gnl|PID|d1015405 (D90746) Outer membrane protein FhuE precursor  
[Escherichia coli] >gi|1787344 (AE000210) outer-membrane receptor for Fe(III)-  
coprogen, Fe(III)-ferrioxamine B and Fe(III)-rhodotruclic acid precursor  
[Escherichia coli] Length = 729  
Score = 332 bits (843), Expect = 3e-90  
60 Identities = 228/717 (31%), Positives = 350/717 (48%), Gaps = 60/717 (8%)

Query: 38 TITVTADRTASSN--DGYTVSGTHTPFGLPMTLREIPQSVSVITSQQMRDQNIKTLDRAL 95  
T+ V TA + + Y+V+ T + MT R+IPQSV++++ Q+M DQ ++TL +  
65 Sbjct: 43 TVIVEGSATAPDDGENDYSVTSTSGATKMQMTQORDIPQSVTIVSQRMEDQQLQTLGEVM 102

Query: 96 LQATGTSRQIYGS DRAGYNYLFARGSR IANYQINGIP-----VADALADTGNANTAA 147  
G S Y ++RG +I NY ++GIP + DAL+D A  
Sbjct: 103 ENTLGISKSQADSRLALY---YSRGFQIDNYMVDGIPTYFESRWNLGALSDM-----AL 154

Query: 148 YERVEVVRGVAGLPDGTGEP SATVNLVRKHPTKPLF-EVRAEAGNRKHFG LGADVSGSL 206  
 +ERVEVVRG GL GTG PSA +N+VRKH T + +V AE G+ AD+ L  
 Sbjct: 155 FERVEVVRGATGLMTGTGNPSAAINMVRKHATSREFKGDVSAEYGSWNKERYVADLQSF 214  
 Query: 207 NAEGLRGLRVSTFGRGDSWRQLERSRDAELYGILEYDIAPQTRVHAGMDYQOAKETADA 266  
 +G +R R+V + DSW S GI++ D+ T + AG +YQ+ +  
 Sbjct: 215 TEDGKIRARIVGGYQNND SWLD RYNSEKTF FSGIVDADLGLD LTL SAGYEYQRIDVNSEPT 274  
 Query: 267 PLSYAVYDSQGYATAFGPKDN PATNWSNSRNLNLFAGIEHFRNQDWKLKAEYDYTRSR 326  
 +++ G + ++ + A +W+ + +F ++ +F W+ ++  
 Sbjct: 275 WGGLPRWNTDGSSNSYDRARSTAPDWAYNDKEINKVFMTLKQFADTWQATLNATHSEVE 334  
 Query: 327 F--RQPYGVAGVLSIDHSTAA--TDLIPGY-----WHADPRTHSA-SMSLTGKYRLEG 374  
 F + Y A V D ++ PG+ W++ R A + G Y LFG  
 Sbjct: 335 FDSKMMYVDAYVNKADGMLVGPYSNYGPGFDYVGGTGWNSGKRKVDALDLFADGSYLEFG 394  
 Query: 375 REHDLIAGINGYKYASNKYGER--SIIPNAIPNAYEFSRTGAYPOPSFAQTIPQYDTRR 432  
 R+H+L+ G Y +N+Y +I P+ I + Y F+ G +PQ Q++ Q Y DT  
 Sbjct: 395 RQHNL MFG-GSYSKQNNRYFSSWANIFPDEIGSFYNFN--GNFPQTDWSPQSLAQDDTTH 451  
 Query: 433 QIGGYLATRFRAADNLSLILGGYRSYRAGSYNSRTQGMTY-VSANRFTPYTGIVFDXXX 491  
 Y ATR AD L LILG RY+ +R + +TY + N TPY G+VFD  
 Sbjct: 452 MKSLYAATRVTLADPLHLILGARYTNWRVDT-----LTYSMEKNHTTPYAGLVFDIND 504  
 Query: 492 XXXXXXXXXXXXFPQLQKDEHGSYLKPV TGNLEADIKGEWLEGRNLNASAAVYRARKNNL 551  
 F PQ +D G YL P+TGNN E +K +W+ RL + A++R ++N+  
 Sbjct: 505 NWSTYASYTSIFQPQND RDSSGKYLAPITGN NYELGLKSDWMNSRLTTT LAIFRIEQDNV 564  
 Query: 552 ATAAGR---DQSGNTYYRAANQAKTHGWEIEVGGRITPEWQIQAGYSQSKPRDQDGSRLN 608  
 A + G +G T Y+A + + G E E+ G IT WQ+ G ++ D +G+ +N  
 Sbjct: 565 AQSTGTPIPGSNGETAYKAVDGT VSKGVEFELNGAITDNWQLTFGATRYIAEDNEGNAVN 624  
 Query: 609 PDSVPERSFKLFTAYHLAPEAPSGRTIGAGVRRQGETHTDPAALRIPNPAKARAVANSR 668  
 P ++P + K+FT+Y L P P T+G GV Q +TD P RA  
 Sbjct: 625 P-NLPRTTVKMFYSYRL-PVMPE-LTVGGGVNWQNRVYTDTV-----TPYGT FRA----E 672  
 Query: 669 QKAYAVADIMARYRFPNRTLSLNVNDLNFKNHYRTQPD RH-SYGALRTVNAAFTYRF 724  
 Q +YA+ D+ RY+ L NV+NLF+K Y T + YG R + TY+F  
 Sbjct: 673 QGSYALVDLFTRYQVTKNFSLQGNVNNLFDKTYDTNVEGSIVYGT PRNFSITGT YQF 729

Based on this analysis, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF23-1 (77.5kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 15A shows the results of affinity purification of the His-fusion protein, and Figure 15B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 15C) and for ELISA (positive result). These experiments confirm that ORF23-1 is a surface-exposed protein, and that it is a useful immunogen.

## Example 80

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 673>:

```

1  ATGCGCACGG CAGTGGTTT GCTGTTGATC ATGCCGATGG CGGCTTCGTC
51  GGCAATGATG CCGGAAATGG TGTGCGCGGG CGTGTGCGCG GGAACGGCAA
101 TCATATCCAA GCCGACCGAA CAAACGGCGG TCATGGCTTC GAGTTTGTC
  
```

151 AGCGTCAGcA CGCCTGCTTC GGCGgcGgCa ATCATACCTT CGTCTTCGGA  
 201 AACGGGGATA AACGcGCCAC TCAAACCCCC GACCGCGCTG GAAGCCATCA  
 251 TGCCGCCTTT TTTCACGGCA TCGTTCAGCA ATGCCAAAGC TGCTGTTGTG  
 301 CCGTGCGTAC CGCAGACGCT CAAGCCCAT TnTTCAAGAA TCGGTGCCAC  
 351 TnAGTCGCCG ACGGGG..

This corresponds to the amino acid sequence <SEQ ID 674; ORF24>:

1 MRTAVVLLLI MPMAASSAMM PEMVCAGVSP GTAIISKPTE QTAVMASSLS  
 51 SVSTPASAAA IIPSSSETGI NAPLKPPTAL EAIMPPFFTA SFSNAKAHV  
 101 PCVPQTLKPI XSRMRATXSP TG..

Further work revealed the complete nucleotide sequence <SEQ ID 675>:

1 ATGCGCACGG CAGTGGTTTT GCTGTTGATC ATGCCGATGG CGGCTTCGTC  
 51 GGCAATGATG CCGGAAATGG TGTGCGCGGG CGTGTGCGCG GGAACGGCAA  
 101 TCATATCCAA GCGGACCGAA CAAACGGCGG TCATGGCTTC GAGTTTGTCC  
 151 AGCGTCAGCA CGCCTGCTTC GGCGGCGGCA ATCATACCTT CGTCTTCGGA  
 201 AACGGGGATA AACGCGCCAC TCAAACCCCC GACCGCGCTG GAAGCCATCA  
 251 TGCCGCCTTT TTTCACGGCA TCGTTCAGCA ATGCCAAAGC TGCTGTTGTG  
 301 CCGTGCGTAC CGCAGACGCT CAAGCCCAT TCTTCAAGAA TCGGTGCCAC  
 351 TGAGTCGCCG ACGGCGGGGG TCGGCGCCAG CGACAAGTCG AGAATACCAA  
 401 ACGGATATT CAGCATTTTT GAGGCTTCGC GGCGGATGAG TTCGCCCACG  
 451 CGGGTAATTT TGAAGCAGT TTTCTTCACT ACTTCCGCAA CTTCGGTCAA  
 501 TGTCGTTGCA TCTGAATTTT CCAACGCGGC TTTTACGACA CCTGGGCCGG  
 551 ATACGCGGAC ATTGATAACG GCATCCGCTT CGCCCGAACC ATGAAACGCG  
 601 CCCGCCATAA ACGGGTTGTC TTCCACCGCG TTGCAGAACA CGACAATTTT  
 651 AGCGCAGCCG AAACCTTCGG GCGTGATTTC CGCCGTGCGT TTGACGGTTT  
 701 CGCCGCGCAG CTTGACCGCA TCCATATTGA TACCGGCACG CGTACTGCCG  
 751 ATATTGATGG AGCTGCACAC AATATCGGTA GTCTTCATCG CTTGGGAAT  
 801 GGAGCGGATT AACACCTCAT CCGAAGGCGA CATCCCTTTT TGCACCAACG  
 851 CGGAAAACCC GCGGATAAAA GACACACCGA TGGCTTTGGC AGCTTTATCC  
 901 AAAGTTTGCG CCACGCTGAC GTAA

This corresponds to the amino acid sequence <SEQ ID 676; ORF24-1>:

1 MRTAVVLLLI MPMAASSAMM PEMVCAGVSP GTAIISKPTE QTAVMASSLS  
 51 SVSTPASAAA IIPSSSETGI NAPLKPPTAL EAIMPPFFTA SFSNAKAHV  
 101 PCVPQTLKPI SSRMRATESP TAGVGASDKS RIPNGIFSIF EASRPMSSPT  
 151 RVILKAVFFT TSATSVNVVA SEFSNAAFIT PGPDPTLIT ASASPEP\*NA  
 201 PAINGLSSTA LQNTTILAQP KPSGVISAVR LTVSPASLTA SILIPARVLP  
 251 ILMELHTISV VFIAISGMERI NTSSEGDIPF CTNAEKPPIK DTPMALAALS  
 301 KVCATLT\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF24 shows 96.4% identity over a 307 aa overlap with an ORF (ORF24a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
orf24a.pep	MRTAVVLLL	IIMPMAASSAMM	PEMVCAGVSP	GTAIISKPTE	QTAVIASSLS	SNVSTPASAAA	
45 orf24	MRTAVVLLL	IIMPMAASSAMM	PEMVCAGVSP	GTAIISKPTE	QTAVMASSLS	SVSTPASAAA	
	10	20	30	40	50	60	
		70	80	90	100	110	120
orf24a.pep	IIPSSSXTG	INAPLKPPTALE	AIMPPFFTA	SFSNAKAHV	PCVPQTLKPI	SSRMRATESP	
50 orf24	IIPSSSETG	INAPLKPPTALE	AIMPPFFTA	SFSNAKAHV	PCVPQTLKPI	SSRMRATESP	
	70	80	90	100	110	120	
		130	140	150	160	170	180
orf24a.pep	TAGVGASDKS	RIPNGIFSIFE	ASRPMSSPTR	VILKAVFFTS	ATSATSVNVV	ASEFSNAAFIT	
orf24	TAGVGASDKS	RIPNGIFSIFE	ASRPMSSPTR	VILKAVFFTS	ATSATSVNVV	ASEFSNAAFIT	

		130	140	150	160	170	180
		190	200	210	220	230	240
5	orf24a.pep	PGPDTPTLITASASPEPXNAPAI XGLSSXALQNTTILA QPKPSSVISXVRLMVSPASLTA					
	orf24	PGPDTPTLITASASPEPXNAPAI NGLSSTALQNTTILA QPKPSGVISAVRLTVSPASLTA					
		190	200	210	220	230	240
10	orf24a.pep	250	260	270	280	290	300
	orf24	SILIPARVLPILMELHTISVVF IASGMERXNTSSEGDIPFCTSA EKPPIKDTPMALAALS					
		250	260	270	280	290	300
15	orf24a.pep	KVCATLTX					
	orf24	KVCATLTX					

The complete length ORF24a nucleotide sequence <SEQ ID 677> is:

20	1	ATGCGCACGG	CAGTGGT	TTTT	GCTGTTGATC	ATGCCGATGG	CGGCTTCGTC
	51	GGCAATGATG	CCGGAAATGG	TGTGCGCGGG	TGTGTCGCGG	GGAACGGCAA	
	101	TCATATCCAA	NCCGACCGAA	CAAACGGCGG	TCATCGCTTC	GAGTTATCC	
	151	AACGTACGCA	CGCCTGCTTC	GGCGGCGGCA	ATCATACCTT	CGTCTCGGA	
25	201	NACGGGGATA	AACGCGCCAC	TCAAACCGCC	AACCGCGCTC	GAAGCCATCA	
	251	TGCCGCCCTT	TTTACCGGCA	TCGTTAGCA	ATGCCAAAGC	TGCTGTTGTG	
	301	CCGTGCGTAC	CGCAGACGCT	CAAACCCATT	TCTTCAAGAA	TGCGCGCCAC	
	351	CGAGTCGCGG	ACGGCAGGGG	TGGTGCCAG	CGACAAGTCG	AGAATACCAA	
	401	ACGGGATATT	CAGCATTTTT	GAGGCTTCGC	GGCCGATGAG	TTCGCCACG	
	451	CGGGTAATTT	TGAAGCGGT	TTTCTTACA	ACTTCGGCAA	CTTCGGTCAA	
30	501	TGTCGTTGCA	TCCGAATTTT	CCAACGCGGC	TTTACGACA	CCCGGGCCGG	
	551	ATACGCCGAC	ATTAATCACA	GCATCCGCTT	CGCCTGAGCC	GTGAAACGGC	
	601	CCCGCCATAN	ACGGGTTGTC	TTCCNCCGCG	TTGCAGAACA	CGACGATTTT	
	651	GGCGCAGCGG	AAACCTTCTA	GTGTGATTTT	ANCCGTGCGT	TTGATGGTTT	
	701	CGCCCGCCAG	TCTGACCGCG	TCCATATTGA	TACCGGCGCG	CGTACTGCCG	
35	751	ATATTGATGG	AGCTGCACAC	GATATCAGTA	GTCTTCATCG	CTTCGGGAAT	
	801	GGAACGGATN	AACACCTCGT	CAGAAGGCGA	CATACCTTTT	TGCACGAGCG	
	851	CGGAAAAGCC	GCCAATAAAA	GACACGCCGA	TGGCTTTGGC	AGCCTTATCC	
	901	AAAGTTTGCG	CCACGCTGAC	GTA			

This encodes a protein having amino acid sequence <SEQ ID 678>:

40	1	MRTAVVLLLI	MPMAASSAMM	PEMVCAGVSP	GTAII SXPTE	QTAVIASSLS
	51	NVSTPASAAA	IIPSSSXTGI	NAPLKPPTAL	EAIMPPFFTA	SFSNAKAADV
	101	PCVPQTLKPI	SSRMRATESP	TAGVGASDKS	RIPNGIFSIF	EASRPMSSPT
	151	RVILKAVFFT	TSATSVNVVA	SEFSNAFTT	PGPDTPTLIT	ASASPEP*NA
45	201	PAIXGLSSXA	LQNTTILAQP	KPSSVISXVR	LMVSPASLTA	SILIPARVLP
	251	ILMELHTISV	VFIASGMERX	NTSSEGDIPF	CTSAEKPIK	DTPMALAALS
	301	KVCATLT*				

It should be noted that this protein includes a stop codon at position 198.

ORF24a and ORF24-1 show 96.4% identity in 307 aa overlap:

50	orf24a.pep	10	20	30	40	50	60
	orf24-1	MRTAVVLLLIMPMAASSAMMPEMVCAGVSPGTAII SXPTEQTAVIASSLSNVSTPASAAA					
		10	20	30	40	50	60
55	orf24a.pep	70	80	90	100	110	120
	orf24-1	IIPSSSXTGINAPLKPPTALEAIMPPFFTASFNSAKAAVPCVPQTLKPISSRMRATESP					
		70	80	90	100	110	120
60		130	140	150	160	170	180

5	orf24a.pep	TAGVGASDKSRIPNGIFSIFEASRPMSSPTRVILKAVFFTTTSATSVNVVASEFSNAAFTT
	orf24-1	TAGVGASDKSRIPNGIFSIFEASRPMSSPTRVILKAVFFTTTSATSVNVVASEFSNAAFTT
10	orf24a.pep	PGPDTPTLITASASPEPXNAPAIKGLSSXALQNTTILAQPKPSSVISXVRLMVSPASLTA
	orf24-1	PGPDTPTLITASASPEPXNAPAIKGLSSXALQNTTILAQPKPSSVISXVRLMVSPASLTA
15	orf24a.pep	SILIPARVLPILMELHTISVVFIASGMRKNTSSEGDIPFCTSAEKKPIKDTMALAALS
	orf24-1	SILIPARVLPILMELHTISVVFIASGMRKNTSSEGDIPFCTSAEKKPIKDTMALAALS
20	orf24a.pep	KVCATLTX
	orf24-1	KVCATLTX

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF24 shows 96.7% identity over a 121 aa overlap with a predicted ORF (ORF24ng) from

### *N.gonorrhoeae*:

25	orf24.pep	MRTAVVLLLIMPMAASSAMMPMVCAGVSPGTAIISKPTAQAVMASSLSVSTPASAAA	60
	orf24ng	MRTAVVLLLIMPMAASSAMMPMVCAGVSPGTAIMSKPTAQAVMASSLSVSTPASAAA	60
30	orf24.pep	IIPSSSETGINAPLKPPTALEAIMPPFFTASFNAKAAVPCVPQTLKPIXRMRATXSP	120
	orf24ng	IIPSSSETGINAPLKPPTALEAIMPPFFTASFNAKAAVPCVPQTLKPIXRMRATESP	120
35	orf24.pep	TG	122
	orf24ng	TG	122
	orf24ng	TAGVGASDKSRIPNGIFSIFEASRPMSSPTRVILKAVFFTTTSATSVRLTASEFSSAALT	180

The complete length ORF24ng nucleotide sequence <SEQ ID 679> is:

40	1	ATGCGCACGG	CGGTGGTTTT	GCTGTTGATC	ATGCCGATGG	CGGCTTCGTC
	51	GGCGATGATG	CCGGAATGG	TGTGCGCGGG	CGTGTGCGCG	GGAACGGCAA
45	101	TCATGTCCAA	ACCAACGGAG	CAGACGGCGG	TCATGGCTTC	GAGTTTGTCC
	151	AGCGTCAACA	CGCTGCGCTC	GGCGGCGGCA	ATCATACCTT	CGTCTCGGA
50	201	AACGGGGATA	AACGCGCCGC	TCAAACCGCC	GACCGCGCTG	GAAGCCATCA
	251	TGCCGCGCTT	TTTCACGGCA	TCGTTAGCA	ATGCCAAAGC	TGCTGTGTG
55	301	CCGTGCGTAC	CGCAGACGCT	CAAGCCCAT	TCTTCAAGAA	TGCGCGCCAC
	351	CGAGTCGCGG	ACGGCGGGGG	TCGGTGCCAG	CGACAAATCG	AGAATGCCGA
60	401	ACGGGATATT	CAGCATTTTT	GAGGCTTCGC	GACCGATGAG	TTCGCCACG
	451	CGGGTGATTT	TGAAAGCGGT	TTTCTTCACG	ACTTCGGCGA	CCTCGGTAC
65	501	GCTGACCGCG	TCCGAATTTT	CCAGCGCGGC	TTTGACCACG	CCTGGACCGG
	551	ATACGCCGAC	ATTAATCACA	GCATCCGCTT	CGCCGAGCC	GTGGAACGCA
70	601	CCCGCCATAA	ACGGATTGTC	TTCCACCGCG	TTGCAGAACA	CGACGATTTT
	651	GGCGCAGCCG	AAACCTTCGG	GTGTGATTTC	AGCCGTGCGT	TGATGGTTT
75	701	CGCCTGCCAG	CTTGACCGCA	TCCATATTGA	TACCGGCACG	CGTGCTGCCG
	751	ATATTGATGG	AGCTGCACAC	GATATCGGTA	GTTTTCATCG	CTTCGGGAAC
80	801	GGAACGGATC	AACACCTCAT	CCGAAGCGCA	CATACCTTTT	TGCACCGCG
	851	CGGAAAAGCC	GCCGATAAAG	GACACGCCGA	TGGCTTTGGC	TGCCTTGTC
	901	AAAGTCTGCG	CCACGCTGAC	ATAA		

This encodes a protein having amino acid sequence <SEQ ID 680>:

60	1	MRTAVVLLLI	MPMAASSAMM	PEMVCAGVSP	GTAIMSKPTE	QTAVMASSLS
	51	SVNTPASAAA	IIPSSSETGI	NAPLKPPTAL	EAIMPPFFTA	SFSNAKAAV
65	101	PCVPQTLKPI	SSRMATESP	TAGVGASDKS	RMNGIFSIF	EASRPMSSPT
	151	RVILKAVFFT	TSATSVRLTA	SEFSSAALT	PGPDTPTLIT	ASASPEPWNA
	201	PAINGLSSTA	LQNTTILAQP	KPSGVISAVR	LMVSPASLTA	SILIPARVLP

251 ILMELHTISV VFIASGTERI NTSSEGDIPF CTSAEKPPIK DTPMALAALS  
 301 KVCATLT\*

ORF24ng and ORF24-1 show 96.1% identity in 307 aa overlap:

		10	20	30	40	50	60
5	orf24-1.pep	MRTAVVLL	LIMPMAASSAM	PEMVCAGVSPG	TAIISKPT	EQTAVMASSL	SSVSTPASAAA
	orf24ng	MRTAVVLL	LIMPMAASSAM	PEMVCAGVSPG	TAIMSKPT	EQTAVMASSL	SSVNTPASAAA
		10	20	30	40	50	60
10	orf24-1.pep	IIPSSSET	GINAPLKPPTA	LEAIMPPFFT	ASFSNAKA	AAVPCVPQ	TLKPISSRM
	orf24ng	IIPSSSET	GINAPLKPPTA	LEAIMPPFFT	ASFSNAKA	AAVPCVPQ	TLKPISSRM
		70	80	90	100	110	120
15	orf24-1.pep	TAGVGASD	KSRIPNGIF	SIFEASRPM	SSPTRVIL	KAVFFTT	SATSVNVV
	orf24ng	TAGVGASD	KSRMPNGIF	SIFEASRPM	SSPTRVIL	KAVFFTT	SATSVRLT
		130	140	150	160	170	180
20	orf24-1.pep	PGPDTPTL	ITASASPEP	XNAPAINGL	SSTALQNT	TILAQPKP	SGVISAVR
	orf24ng	PGPDTPTL	ITASASPEP	WNAPAINGL	SSTALQNT	TILAQPKP	SGVISAVR
		190	200	210	220	230	240
25	orf24-1.pep	SILIPARV	LPILMELHT	ISVVFIASG	MERINTS	SEGDIPF	CTNAEKPP
	orf24ng	SILIPARV	LPILMELHT	ISVVFIASG	TERINTS	SEGDIPF	CTSAEKPP
		250	260	270	280	290	300
30	orf24-1.pep	KVCATLT	X				
	orf24ng	KVCATLT	X				
35	orf24-1.pep	KVCATLT	X				
	orf24ng	KVCATLT	X				

Based on this analysis, including the presence of a putative leader sequence (first 18 aa – double-underlined) and putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 81

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 681>:

45	1	..ACCGACGTGC	AAAAAGAGTT	GGTCGGCGAA	CAACGCAAGT	GGGCGCAGGA
	51	AAAAATCAGC	AACTGCCGAC	AAGCCGCCGC	GCAGGCAGAC	CGGCAGGAAT
	101	ACGCCGAATA	CCTCAAGCTG	CAATGCGACA	CGCGGATGAC	GCGCGAACGG
	151	ATACAGTATC	TTGCGGCTA	TTCCATCGAT	TAG	

This corresponds to the amino acid sequence <SEQ ID 682; ORF25>:

50	1	..TDVQKELVGE	QRKWAQEKIS	NCRQAAQAD	RQEYAEYLKL	QCDTRMTRE
	51	IQYLRGYSID	*			

Further work revealed the complete nucleotide sequence <SEQ ID 683>:

	1	ATGTATCGGA	AACTCATTGC	GCTGCCGTTT	GCCCTGCTGC	TTGCCGCTTG
	51	CGGCAGGGAA	GAACCGCCCA	AGGCATTGGA	ATGCGCCAAC	CCCGCCGTGT

10

15

101	TGCAAGGCAT	ACGCGGCAAT	ATTCAAGAAA	CGCTACGCA	GGAAGCGCGT
151	TCTTTCGCGC	GCGAAGACGG	CAGGCAGTTT	GTCGATGCCG	ACAAAATTAT
201	CGCCCGCCGC	TACGTTTGG	CGCTTTCTTT	GGAACACGCT	TCCGGAACCG
251	AGGAAGCGCG	GCGCAGCTT	TGTATCGCCG	ATTTGAACAT	TACGGTGCCG
301	TCTGAAACGC	TGTGCCATGC	CAAGGCAAAC	AGCCCCCTGT	TGTACGGGGA
351	AACTGCTTTG	TCGGATATTG	TGCGCGAGAA	GACGGGCGCG	AATGTCCGAGT
401	TTAAAGACGG	CGTATTGACG	GACGCCGTCC	GCTTCTGTCC	CGTCAACAGT
451	GGTCAGACGG	CATTTGTCSA	CAACACGGTC	GGTATGCGCG	CGCAAACGCT
501	GTCTGCCGCG	CTGCTGCCTT	ACGGCGTGAA	GAGCATCGTG	ATGATAGACG
551	GCAAGCGCGT	GAAAAAGAA	GACGCGGTCA	GGATTTTGAG	CGGAAAAGCC
601	CGTGAAGAAG	AACCGTCCAA	ACCCACGCCC	GAAACATATT	TGGAACACAA
651	TGCCCGCCGC	GGCAGTTGCG	CGGTACCCCA	AGCCGCGAGAA	GGCGCGCCCG
701	AACCGGAAAT	CCTGCATCCT	GACGACGGCG	AGCGTGCCCA	TACCGTTACC
751	GTATCACGGG	GCGAAGTGGA	AGAGGCGCGC	GTACAAAACC	AGCTGTCCGA
801	ATCCGAATAAT	ACCAAATCTT	GGGGAGGAGT	CGATACCGAC	TGCAAAAAGG
851	AGTTGGTTCG	CGAACAACGC	AAGTGGGCGC	AGGAAAAAAT	CAGCAACTGC
901	CGACAGATCG	CCGCGCAGGC	AGACCGGCAG	GAATACGCCG	AATACCTCAA
951	GCTGCAATGC	GACACGCGGA	TGACGCGCGA	ACGGATACAG	TATCTTCCGG
1001	GCTATTCCAT	GCATTAG			

20 This corresponds to the amino acid sequence <SEQ ID 684; ORF25-1>:

25

```

1 MYRKLIALPF ALLAACGRE EPPKALECAN PAVLQGIRGN IQETLTQEAR
51 SFAREDKGRQF VDADKIITAA YGLFASLEHA SETVEGGRTF CIADLNTIVP
101 SETLADAKAN SPLLYETAL SDIVRQKTGG NVEFGKGVLT AAVRLFPVKD
151 GQAFVDNVTV GMAAQETLSAA LLPYGVKSIV MIDGKAVKKE DAVRILSGKA
201 REEFPKPTP EDILHNAAG GDAGVQPAAE GAPEPEILHP DDGERADTVT
251 VSRGEVEEAR VQNRASEI TKLWGLDSTD VQKELVGEQR KWAQEKISNC
301 ROAAAOADRO EYAEYLKQOC DTRMTRERIQ YLRGYSID*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

30 ORF25 shows 98.3% identity over a 60aa overlap with an ORF (ORF25a) from strain A of *N. meningitidis*:

35

40

```

                                10      20      30
orf25.pep                      TDVQKELVGEQRKWAEKISNCRQAAAQAD
                                ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
orf25a      VTVSRGEVEEARVQNQRAESEITKLWGGLD TDVQKELVGEQRKWAEKISNCRQAAAQAD
            250      260      270      280      290      300

                                40      50      60
orf25.pep      RQEYAEYKLQCDTRMTRERIQYLRGYSIDX
                ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
orf25a      RQEYAEYKLQCDTRMTRERIQYLRGYSIDX
            310      320      330

```

The complete length ORF25a nucleotide sequence <SEQ ID 685> is:

45

50

55

60

1	ATGTATCGGA	AACCTATTGC	GCTGCCGTTT	GCCCTGCTGC	TTGCCGCTTG
51	CGGCAGGGAA	GAACGCCCCA	AGGCATTGGA	ATGCGCCAAC	CCCAGCCTGT
101	TCTCAANGCAT	ACGCNGCAAT	ATTCAGGAAA	CGCTCACGCA	GGAGCGCTGT
151	TCATTTCGCG	CGCAAGACNG	CANGCAGTTT	TCGATGCGCG	ACNAAATTAT
201	CGCCGCGCGC	TANGNTNNNG	NGNTNTCTTT	GGAACACGCT	TCGGAAACGC
251	AGGAAGGCGG	GGGCACGTTT	TGNTNTCGCG	ATTTGAACAT	TACCGTGCCG
301	TCTGAAACGC	TTGCCGATGC	CAAGGCAAAC	AGCCCCCTGC	TGTACGGGGA
351	AACCGCTTTT	TCGGATATTG	TGCGGCAGAA	GACGGGCGCG	AATGTCGAGT
401	TTAAAGACGG	CGTATTGACG	GCAGCGCTCC	GCTTCTTACC	CGTAAAGAC
451	GGTCAGANGC	CATTTGTGCA	CAACACGGTC	GGTATGCGCG	GCCTAAACGCT
501	GTCTGCCGCG	TTGCTGCCTT	ACGGCGTGAA	GAGCATCGTG	ATGATAGACG
551	GCAAGGCGGT	AAAAAAGAA	GACCGCGTCA	GGATTNTGAG	CNGNAAGCC
601	CGTGAANAAG	AACCTGCTCA	ANCNNNGCC	GAAGACATTT	TGGAACATTA
651	TGCCCGCGGA	GGGGATGCAG	ACGTACCCCA	AGCCGAGAA	GACGCGCCCG
701	AACCGGAAAT	CCTGCATCTT	GACGACGGCG	AGTCGTCCGA	TACCGTTACC
751	GTATCACAGG	GCGAAGTGGA	AGAGGCGCGN	GTACAAAACC	AGCGTGCGGA
801	ATCCGAAATT	ACCAAATTTT	GGGGAGGACT	CGATACCGAC	TGTCAAAAAG



851 AGTTGGTCGG CGAANAACGC AAGTGGGCGC AGGAAAAAAT CAGCAACTGC  
 901 CGACAAGCCG CCGCGCAGGC AGACCGGCAG GAATACGCCG AATACCTCAA  
 951 GCTGCAATGC GACACGCGGA TGACGCGCGA ACGGATACAG TATCTTCGCG  
 1001 GCTATTCCAT CGATTAG

5 This encodes a protein having amino acid sequence <SEQ ID 686>:

1 MYRKLIALPF ALLLAACGRE EPPKALECAN PAVLQXIRXN IQETLTQEAR  
 51 SFAREDXXQF VDADXIIAAA XXXXSLEHA SETQEGGRTF CXADLNITVP  
 101 SETLADAKAN SPLLYGETAL SDIVRQKTGG NVEFKDGLT AAVRFLPVKD  
 151 QXAFVDNTV GMAAQTLSAA LLPYGVKSIV MIDGKAVKKE DAVRIXSXXA  
 201 REXEPSKXXP EDILEHNAAG GDADVQAGE DAPEPEILHP DDGERADTVT  
 251 VSRGEVEEAR VQNQRAESEI TKLWGGLDTD VQKELVGEXR KWAQEKISNC  
 301 RQAAQADRQ EYAEYLKLCQ DTRMTREIRI YLRGYSID\*

ORF25a and ORF25-1 show 93.5% identity in 338 aa overlap:

15	orf25a.pep	10	20	30	40	50	60
		MYRKLIALPFALLLAACGREPPKALECANPAVLQXIRXNIQETLTQEARSFAREDXXQF					
	orf25-1	MYRKLIALPFALLLAACGREPPKALECANPAVLQGIQIRNIQETLTQEARSFAREDGRQF					
20		10	20	30	40	50	60
	orf25a.pep	70	80	90	100	110	120
		VDADXIIAAAXXXXSLEHASETQEGGRTFCXADLNITVPSETLADAKANSPLLYGETAL					
	orf25-1	VDADXIIAAAYGLAFSLEHASETQEGGRTFCIADLNITVPSETLADAKANSPLLYGETAL					
25		70	80	90	100	110	120
	orf25a.pep	130	140	150	160	170	180
		SDIVRQKTGGNVEFKDGLTAAVRFLPVKDGQXAFVDNTVGMAAQTLSAALLPYGVKSIV					
	orf25-1	SDIVRQKTGGNVEFKDGLTAAVRFLPVKDGQTAFVDNTVGMAAQTLSAALLPYGVKSIV					
30		130	140	150	160	170	180
	orf25a.pep	190	200	210	220	230	240
		MIDGKAVKKEDAVRIXSXXAREXEPSKXXPEDILEHNAAGGDADVQAGEDAPEPEILHP					
	orf25-1	MIDGKAVKKEDAVRILSGKAREEPEPSKPTPEDILEHNAAGGDAGVQAAEGAPEPEILHP					
35		190	200	210	220	230	240
	orf25a.pep	250	260	270	280	290	300
		DDGERADTVTVSRGEVEEARVQNQRAESEITKLWGGLDTDVQKELVGEXRKWAQEKISNC					
	orf25-1	DDGERADTVTVSRGEVEEARVQNQRAESEITKLWGGLDTDVQKELVGEXRKWAQEKISNC					
40		250	260	270	280	290	300
	orf25a.pep	310	320	330	339		
		RQAAQADRQYAEYLKLCQDTRMTREIRIYLRGYSIDX					
	orf25-1	RQAAQADRQYAEYLKLCQDTRMTREIRIYLRGYSIDX					
45		310	320	330			

50 Homology with a predicted ORF from *N.gonorrhoeae*

ORF25 shows 100% identity over a 60aa overlap with a predicted ORF (ORF25ng) from *N.gonorrhoeae*:

55	orf25.pep	TDVQKELVGEXRKWAQEKISNCRQAAQAD	30
	orf25ng	VTVSRGEVEEARVQNQRAESEITKLWGGLDTDVQKELVGEXRKWAQEKISNCRQAAQAD	308
	orf25.pep	RQEYAEYLKLCQDTRMTREIRIYLRGYSID	60
	orf25ng	RQEYAEYLKLCQDTRMTREIRIYLRGYSID	338

60 The complete length ORF25ng nucleotide sequence <SEQ ID 687> is:

5  
10  
15  
20

```

1  ATGTATCGGA AACTCATTGC GCTGCCGTTT GCCCTGCTGC TTGCAGCGTG
51 CGGCAGGGAA GAACCGCCCA AGGCGTTGGA ATGCGCCAAC CCCGCCGTGT
101 TGCAGGACAT ACGCGGCAGT ATTCAGGAAA CGCTACGCA GGAAGCGCGT
151 TCTTTCGCGC GCGAAGACGG CAGGCAGTTT GTCGATGCCG ACAAATATAT
201 CGCCGCCGCC TACGGTTTGG CGTTTTCTTT GGAACACGCT TCGGAAACGC
251 AGGAAGGCGG GCGCACGTTC TGTATCGCCG ATTTGAACAT TACCGTGCCG
301 TCTGAAACGC TTGCCGATGC CGAGGCAAAC AGCCCTCTGC TGTATGGGGA
351 AACGTCTTTG GCAGACATCG TGCAGCAGAA GACGGGCGGC AATGTCGAGT
401 TTAAGACGG CGTATTGACG GCAGCCGTCC GCTTCTGCCC CGCAAAGAC
451 GCTCGGACGG CATTATCGA CAACACGGTC GGTATGGCGA CGCAAACGCT
501 GTCTGCCCGG TTGCTGCCTT ACGGCGTGAA GAGCATCGTG ATGATAGACG
551 GCAAGGCGGT GACAAAAGAA GACGCGGTCA GGGTTTGTAG CGGCAAAGCC
601 CGTGAAGAAG AACCGTCCAA ACCCACCCCC GAAGACATTT TGGAAACAAA
651 TGCCGCCGGC GCGATGCGG GCGTACCCCA AGCCGCAGAA GGCGCACCCG
701 AACCCGAAAT CCTGCATCCC GACGACGTCG AGCGTGCCGA TACCGTTACC
751 GTATCACGGG GCGAAGTGGA AGAGGCGCGC GTACAAAACC AACGTGCGGA
801 ATCCGAAATT ACCAACTTT GGGGAGGACT CGATACCGAC GTGCAAAAAG
851 AGTTGGTCGG CGAACAGCGC AAGTGGGCGC AGGAAAAAAT CAGGcaactgc
901 cgACAAGCCG CCGCGCAGGC AGACGGGCAG GAATACGCCG AATACCTCAA
951 GCTCCAATGC GACACGCGGA TGACGCGCGA ACggaTACAG TATCTTCGCG
1001 GCTATTCCAT CGATTAG

```

This encodes a protein having amino acid sequence <SEQ ID 688>:

25

```

1  MYRKLIALPF ALLLAACGRE EPPKALECAN PAVLQDIRGS IQETLTQEAR
51 SFAREDGRQF VDADKIIAAA YGLAFSLEHA SETQEGGRTF CIADLNITVP
101 SETLADAEAN SPILYGETSL ADIVQKQTGG NVEFKDGLVT AAVRFLPAKD
151 ARTAFIDNTV GMATQTLASA LLPYGVKSIV MIDGKAVTKE DAVRVLGSKA
201 REEESKPTP EDILEHNAAG GDAGVPQAAE GAPEPEILHP DDVERADTVT
251 VSRGEVEEAR VQNQRAESEI TKLWGGLDLD VQKELVGEQR KWAQEKISNC
301 RQAAAQADRQ EYAEYKLQCD TRMTRERIQ YLRGYSID*

```

30 ORF25ng and ORF25-1 show 95.9% identity in 338 aa overlap:

35  
40  
45  
50  
55  
60  
65

	10	20	30	40	50	60
orf25-1.pep	MYRKLIALPFALLLAACGRE	EPPKALECANPAVLQ	GIRGNIQETLTQEAR	SFAREDGRQF		
orf25ng	MYRKLIALPFALLLAACGRE	EPPKALECANPAVLQ	DIRGSIQETLTQEAR	SFAREDGRQF		
	10	20	30	40	50	60
orf25-1.pep	VDADKIIAAAYGLAFSLE	HASETQEGGRTFCIAD	LNITVPSETLADAKAN	SPILYGETAL		
orf25ng	VDADKIIAAAYGLAFSLE	HASETQEGGRTFCIAD	LNITVPSETLADAEAN	SPILYGETSL		
	70	80	90	100	110	120
orf25-1.pep	SDIVRQKTGGNVEFKDGL	TAAVRFLPVKDQTA	FVDNTVGMAAQTL	SAALLPYGVKSIV		
orf25ng	ADIVQKQTGGNVEFKDGL	TAAVRFLPAKDART	AFIDNTVGMATQTL	SAALLPYGVKSIV		
	130	140	150	160	170	180
orf25-1.pep	MIDGKAVKKEDAVRILSG	KAREEESKPTPEDILE	HNAAGGDAGVPQAA	EGAPEPEILHP		
orf25ng	MIDGKAVTKEDAVRVL	SGKAREEESKPTPED	ILEHNAAGGDAGVP	QAAEGAPEPEILHP		
	190	200	210	220	230	240
orf25-1.pep	DDGERADTVTVSRGEVE	EARVQNQRAESEIT	KLWGGLDLDVQKEL	VGEQRKWAQEKISNC		
orf25ng	DDVERADTVTVSRGEVE	EARVQNQRAESEIT	KLWGGLDLDVQKEL	VGEQRKWAQEKISNC		
	250	260	270	280	290	300
orf25-1.pep	RQAAAQADRQEYAEYKL	QCDTRMTRERIQYLR	GYSIDX			
orf25ng	RQAAAQADRQEYAEYKL	QCDTRMTRERIQYLR	GYSIDX			
	310	320	330	339		

Based on this analysis, including the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site (underlined) in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

- 5 ORF25-1 (37kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 16A shows the results of affinity purification of the GST-fusion protein, and Figure 16B shows the results of expression of the His-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 16C), ELISA (positive result), and FACS analysis (Figure 16D). These experiments confirm that ORF25-1 is a surface-exposed protein, and that it is a useful immunogen.

Figure 16E shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF25-1.

### Example 82

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 689>

```

15      1  ATGCAGCTGA  TCGACTATTC  ACATTCATTT  TTCTCGGTTG  TGCCACCCTT
      51  TTTGGCACTG  GCACTTGCCG  TCATTACCCG  CCGCGTACTG  CTGTCTTTAG
     101  GCATCGGTAT  TCTGGwysGC  GTTGCCTTTT  TGGTCGGCGG  CAACCCCGTC
     151  GACGGTCTGA  CACACCTGAA  AGACATGGTC  GTCGGCTTGG  CTGGTGCAGA
     201  CGsyGATTGG  TCGCTGGGCA  AACCAAAAT  CTTGGTTTTT  CkGATACTTT
     251  TGGGTATTTT  TACTTCCCTG  CTGACCTACT  CCGGCAGCAA  T.....
                                     //
     851  .....AC  TTCGCTGGTA
     901  TTCGGCGGCA  CTTGCGGCGT  CTTTGCCGTC  GTTCTCTGCA  CGCTCGGCAC
     951  GATTAAACC  GCCGACTATC  CCAAAGCCGT  TTGGCAGGGT  GCGAAATCTA
    1001  TGTTCGGCGC  AATCGCCATT  TTAATCCTCG  CTGGGCTCAT  CAGTACGGTT
    1051  GTCGGCGAAA  TGCACACCGG  CGATTACCTC  TCCACACTGG  TTGCGGGCAA
    1101  CATCCATCCC  GGCTTCCTGC  CCGTCATCCT  CTCCTGCTC  GCCAGCGTGA
    1151  TGGCGTTTGC  CACAGGCACA  AGCTGGGGGA  CGTTCGGCAT  TATGCTGCCG
    1201  ATTGCCGCCG  CCATGGCGGT  CAAAGTCGAA  CCCGCGCTGA  TTATCCCGTG
    1251  TATGTCCGCA  GTAATGGCGG  GGGCGGTATG  CGGCGACCAC  TGCTCGCCCA
    1301  TTTCCGACAC  GACCATCCTG  TCGTCCACCG  GCGCGCGCTG  CAACCACATC
    1351  GACCACGTTA  CCTCGCAACT  GCCTTACGCC  TTAACCGTTG  CCGCCGCCGC
    1401  CGCATCGGGC  TACCTCGCAT  TGGGTCTGAC  AAAATCCGCG  CTGTTGGGCT
    1451  TTGGCACGAC  AGGCATTGTA  TTGGCGGTGC  TGATTTTCT  GTTGAAAGAT
    1501  AAAAAA..
  
```

This corresponds to the amino acid sequence <SEQ ID 690; ORF26>:

```

40      1  MQLIDYSHSF  FSVVPFLAL  ALAVITRRVL  LSLGIGILXX  VAEFLVGGNPV
     51  DGLTHLKDMV  VGLAWSXDW  SLGKPKILVF  XILLGIFTSL  LTYSGSN...
                                     //
    251  .....TSLV
    301  FGGTCGVFAV  VLCTLGTIKT  ADYPKAVWQG  AKSMFGAIAI  LILAWLISTV
    351  VGEMHTGDYL  STLVAGNIHP  GFLPVILFLL  ASVMAFATGT  SWGTFGIMLP
    401  IAAAMAVKVE  PALIIPCMSA  VMAGAVCGDH  CSPISDTTIL  SSTGARCNIH
    451  DHVTSQLPYA  LTVAAAAASG  YLALGLTKSA  LLGFGTTGIV  LAVLIFLLKD
    501  KK..
  
```

Further work revealed the complete nucleotide sequence <SEQ ID 691>:

1 ATGCAGCTGA TCGACTATTC ACATTCATTT TTCTCGGTTG TGCCACCCTT  
 51 TTTGGCACTG GCACTTGCCG TCATTACCGG CCGCGTACTG CTGTCTTTAG  
 101 GCATCGGTAT TCTGGTCGGC GTTGCCTTTT TGGTCGGCGG CAACCCCGTC  
 151 GACGGTCTGA CACACCTGAA AGACATGGTC GTCGGCTTGG CTTGGTCAGA  
 5 201 CGGCGATTGG TCGCTGGGCA AACCAAAAT CTTGGTTTTT CTGATACTTT  
 251 TGGGTATTTT TACTTCCCTG CTGACCTACT CCGGCAGCAA TCAGGCGTTT  
 301 GCCGACTGGG CAAAACGGCA CATTAAAAAC CGGCGCGGCG CGAAAAATGCT  
 351 GACCGCTGCG CTCGTGTTCG TAACCTTTAT CGACGACTAT TTCCACAGTC  
 401 TCGCCGTCGG TCGGATTGCC CGCCCCGTTA CCGACAAGTT TAAAGTTTCC  
 10 451 CGCACCAAAC TCGCCTACAT CCTCGACTCC ACTGCCGCTC CTATGTGCGT  
 501 GCTGATGCCC GTTTCAGGCT GGGGCGCGTC GATTATCGCC ACGCTTGCCG  
 551 GACTGCTCGT TACCTACAAA ATCACC GAAT ACACGCCGAT GGGGACGTTT  
 601 GTCGCCATGA GCCTGATGAA CTATTACGCA CTGTTTGCCC TGATTATGGT  
 651 GTTCGTGCTC GCATGGTTTT CCTTCGACAT CGGCTCGATG GCACGTTTCG  
 15 701 AACAAAGCCG GTTGAACGAA GCCACGATG AAAGTGGCGT TTCAGACGCT  
 751 ACCAAAGGTC GTGTTTACGC ACTGATTATT CCCGTTTTGG CCTTAATCGC  
 801 CTCACGGTT TCCGCCATGA TCTACACCGG CGCGCAGGCA AGCGAAACCT  
 851 TCAGCATTTT GGGGGCATT GAAAACACGG ACGTAAACAC TTCGCTGGTA  
 901 TTCGCGGCA CTTGCGCGCT CCTTGCCGTC GTTCTCTGCA CGCTCGGCAC  
 20 951 GATTAAACC GCCGACTATC CCAAAGCCGT TTGGCAGGGT GCGAAATCTA  
 1001 TGTTGCGCGC AATCGCCATT TTAATCCTCG CTTGGCTCAT CAGTACGGTT  
 1051 GTCGGCGAAA TGCACACCGG CGATTACCTC TCCACACTGG TTGCGGGCAA  
 1101 CATCCATCCC GGCTTCTGCG CCGTCATCCT CTTCCTGCTC GCCAGCGTGA  
 1151 TGGCGTTTGC CACAGGCACA AGCTGGGGGA CGTTCGGCAT TATGCTGCCG  
 25 1201 ATTGCCGCGG CCATGGCGGT CAAAGTCGAA CCCGCGCTGA TTATCCCGTG  
 1251 TATGTCCGCA GTAATGGCGG GGGCGGTATG CGGCGACCAC TGCTCGCCCA  
 1301 TTTCCGACAC GACCATCCTG TCGTCCACCG GCGCGCGCTG CAACCACATC  
 1351 GACCACGTTA CCTCGCAACT GCCTTACGCC TTAACCGTTG CCGCCGCCGC  
 1401 CGCATCGGGC TACCTCGCAT TGGGTCTGAC AAAATCCGCG CTGTTGGGCT  
 30 1451 TTGGCACGAC AGGCATTGTA TTGGCGGTGC TGATTTTCT GTTGAAAGAT  
 1501 AAAAAACGCG CCAACGCCTG A

This corresponds to the amino acid sequence <SEQ ID 692; ORF26-1>:

1 MQLIDYSHSF FSVVPPFLAL ALAVITRRVL LSLGIGILVG VAFVGGNPV  
 51 DGLTHLKDMV VGLAWSDDGW SLGKPKILVF LILGIFTSL LTYSGSNQAF  
 101 ADWAKRHIKN RRGAKMLTAC LVFVTFIDY FHS LAVGAIA RPVTDKFKVS  
 151 RTKLAYILDS TAAPMCVIMP VSSWGASIIA TLAGLLVTYK ITEYTPMGT  
 201 VAMSLMNYIA LFALIMVFVV AWFSFDIGSM ARFEQAALNE AHDETAVS DA  
 251 TKGRVYALII FVLALIASTV SAMIYTGAQA SETFSILGAF ENT DVNTSLV  
 301 FGGTCGVLA V LCTLTGTTK ADYPAVWQG AKSMFEGAIAI LILAWLISTV  
 351 VGEMHTGDYL STLVAAGNIHP GFLPVILFLL ASVMAFATGT SWGTFGIMLP  
 401 IAAAMAVKVE PALIIPCMSA VMAGAVCGDH CSPISD TTIL SSTGARNHI  
 451 DHVTSQLPYA LTVAAAAASG YLALGLTKSA LLGFETTGI V LAVLIFLLKD  
 501 KKRANA\*

Computer analysis of this amino acid sequence gave the following results:

- 45 Homology with the hypothetical transmembrane protein HI1586 of *H. influenzae* (accession number P44263)  
 ORF26 and HI1586 show 53% and 49% amino acid identity in 97 and 221 aa overlap at the  
 N-terminus and C-terminus, respectively:

Orf26 1 MQLIDYSHSF FSVVPPFLALALAVITRRVXXXXXXXXXXXXVAFVGGNPVDGLTHLKDMV 60  
 M+LID+S S +S+VP LA+ LA+ TRRV L +L V  
 50 HI1586 14 MELIDFSSSVWSIVPALLAIIAIAITRRVLVSLSAGIIIGSLMSDWQIGSAFNVLVKNV 73  
 Orf26 61 VGLAWSDXDWSLGKPKILVFIXLLGIFTSLTYSGSN 97  
 V L ++D + + I++F +LLG+ T+LLT SGSN  
 55 HI1586 74 VSLVYADGEIN-SNMNIVLFLLLLGLVLTALLTVSGSN 109  
 //  
 Orf26 86 IFTSLTYSGS--NTSLVFGGTCGVFAVVLCTL--GTIKTADYPAVWQGA KSMFGXXXX 141  
 +F+ L T+ + TSLV GG C + L + + +Y ++ G KSM G  
 60 HI1586 299 VFSVLGTFENTVVGTSLVVGGFCSIIISTLLIILDRQVSVPYVRSWIVGIKSMGAIAI 358

Orf26 142 XXXXXXXSTVVGEMHTGDYLSLTVAGNIHPGFLPVILFLLASVMAFATGTSWGTFGIMLP 201  
 + +VG+M TG YLS+LV+GNI FLPVILF+L + MAF+TGTSWGTFGIMLP  
 HI1586 359 LFFAWTINKIVGDMQTGKYLSLVSGNIPMQFLPVILFVLGAAMAFSTGTSWGTFGIMLP 418

5 Orf26 202 IAAAMAVKVEPALIIPCMSAVMAGAVCGDHCSPISDTTILSSTGARNHIDHVT SQXXXX 261  
 IAAAMA P L++PC+SAVMAGAVCGDHCSP+SDTTILSSTGA+CNHIDHVT+Q  
 HI1586 419 IAAAMAANAPELLEPCLSAVMAGAVCGDHCSPVSDTTILSSTGAKCNHIDHVTQLPYA 478

10 Orf26 262 XXXXXXXXXXXXXXXXXXXKSALLGFGTTGIVLAVLIFLLKDK 302  
 S L GF T + L V+IF +K +  
 HI1586 479 ATVATATSIGYIVVGFTYSGLAGFAATAVSLVIFAVKKR 519

### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF26 shows 58.2% identity over a 502aa overlap with an ORF (ORF26a) from strain A of *N. meningitidis*:

20 orf26.pep MQLIDYSHSFFSVPPFLALALAVITRRVLLSLGIGILXXVAFVLVGGNPVDGLTHLKDMV  
 orf26a MQLIDYSHSFFSVPPFLALALAVITRRVLLSLGIGILVGVAFVLVGGNPVDGLTHLKDMV

25 orf26.pep VGLAWSDXDWSLGPVKILVFXILLGIFTSLTTYSGSNXX-----  
 orf26a VGLAWSGDGWSLGPVKXLVFLILLGIFTSLTTYSGSNQAFADWAKRHIKNRRGAKMLTAC

30 orf26.pep -----  
 orf26a LVFVTFIDDYFHS LAVGAXARPVTDKFKVSRAKLAYILDSTAAPMCVLMFVSSWGASIIA

35 orf26.pep -----  
 orf26a TLAGLLVYKITEYTPMGT FVAMSLMNYALFALIMVFVVAWFSFDIGSMARFEQAALNE

40 orf26.pep ----- 100 110 -----TSLV  
 orf26a AHDETAVSDGSGR VYALII PVLALIASTVSAMIYTGAQASETF SILGAFENTDVNTSLV

45 orf26.pep 120 130 140 150 160 170  
 FGGTCGVFAVVLCTLTGTIKTADYPKAVWQAKSMFGAIAILILAWLISTVVGEMHTGDYL  
 orf26a FGGTCGVLA VVLCTLTGTIKIADYPKAVWQAKSMFGAIAILILAWLISTVVGEMHTGDYL

50 orf26.pep 180 190 200 210 220 230  
 STL VAGNIHPGFLPVILFLLASVMAFATGTSWGTFGIMLP IAAAMAVKVEPALIIPCMSA  
 orf26a STL VAGNIHPGFLXVILFLLASVMAFATGTSWGTFGIMLP IAAAMAVKVDP SLIIPCMSA

55 orf26.pep 240 250 260 270 280 290  
 VMAGAVCGDHCSPISDTTILSSTGARNHIDHVTSQLPYALTVA AAAASGYLALGLTKSA  
 orf26a VMAGAVCGDHCSPISDTTILSSTGARNHIDHVTSQLPYALTVA AAAASGYLALGLTKSA

60 orf26.pep 300 310  
 LLGFGTTGIVLAVLIFLLKDKK  
 orf26a LLGFGTTGIVLAVLIFLLKDKK

5	1	ATGCAGCTGA	TCGACTATTTC	ACATTCATT	TTCTCGGTTG	TGCCACCCCT
	51	TTTGGGACTG	GCACTTGCCG	TCATTACCCG	CCGCGTACTG	CTGTCTTTAG
	101	GCATCCGGAT	TCTGGTCGGC	GTTGCCCTTT	TGGTCGGCGG	CAACCCCGTC
	151	GACGGTCTGA	CACACCTGAA	AGACATGGTC	TCGCGCTTGG	CTTGGTCAGA
10	201	CGGCGATTGG	TCGCTGGGCA	AACCAAAANT	CTTGGTTTTCT	CTGATACTTT
	251	TGGGTATTTT	TACTTCCCTG	CTGACCTACT	CCGGCAGCAA	TCAGGCGGTT
	301	CGCGACTGGG	CAAAACGGCA	CATTA AAAAC	CGCGCGCGCG	CGAAAAATGCT
	351	GACCGCCTGC	CTCGTGTTCT	TAACCTTTAT	CGACGACTAT	TTCCACAGTC
15	401	TCGCCCTCGG	TGCGNTTGCC	CGCCCCGTTA	CCGACAAGTT	TAAAGTTTTCC
	451	CGCGCCAAAC	TCGCCATCAT	CCTCGACTCC	ACTGCCGCGC	CTATGTGCGT
	501	GCTGATGCCC	GTTTCAAGCT	GGGGCGCGTC	GATTATCGCC	ACGCTTGCCG
	551	GACTGCTCGT	TACCTACAAA	ATCACCGAAT	ACACGCCGAT	GGGGACGTTT
20	601	GTCGCCATGA	GCCTGATGAA	CTATTACGCA	CTGTTTGCCC	TGATTATGGT
	651	GTTTCGTCGT	GCAATGGTCT	CCTTCGACAT	CGGCTCGATG	GACAGTTTCG
	701	AACAAGCCGC	GTTGAACGAA	GCCACGATG	AAACTGCCGT	TTCAGACGGC
	751	AGCTGGGGCA	GGGTTTACGC	ATTGATTATT	CCGGTTTTGG	CCTTAATTCG
25	801	CTCAACGGTT	TCCGCTATGA	TCTAACCCGG	TGCACAGGCA	AGCGAAACCT
	851	TCAGCATTTT	GGGTGCATTT	GAAAACACCG	ACGTGAACAC	TTTCGCTGGTA
	901	FTCGGCGGCA	CTTGCGCGCT	GCTTGCCGTC	GTCTCTGCA	CGCTCGGCAC
	951	GATTA AAAAT	CCCGATTATC	CCAAAGCCGT	TTGGCAGGGT	CGCAAAATCCA
30	1001	TGTTTCGCGC	AATCGCCATT	TTAATCCTTG	CCTGGCTCAT	GAGTACGGTT
	1051	GTCGGCGAAA	TGCACACAGG	CGACTACCTC	TCCACGCTGG	TTGCGGGCAA
	1101	CATCCATCCC	GGCTTCTCTN	CCGTCACTCT	TTTCTCTGCT	GCCAGCGGTA
	1151	TGGCGTTTTC	CACAGGCA	AGCTGGGGGA	CGTTCCGATC	CATGCTCGCG
35	1201	ATTGCCGCGG	CCATGGCGGT	CAAAGTCGAT	CCCTCACTGA	TTATCCCGTG
	1251	TATGTCGCGC	GTGATGCGGG	GGGCGGTATG	CGGCAGCCAC	TGCTCGCCCA
	1301	TTTCCGACAC	GACCATCTCT	TCGTCCACCG	CGCGCGCTG	CAACACATC
	1351	GACCACGTTA	CNTCGCAACT	GCCTTACGCC	TTAACGGTTG	CGCCCGCGCG
	1401	CGCATCGGGN	TACCTCGCAT	TGGGTCTGAC	AAAATCCGCG	CTGTTGGGTT
	1451	TTGGCGANGC	AGGCATTGTA	TGGGCGGTGC	TGATTTTTCT	GTTGAAAGAT
36	1501	AAAAAACCGC	CCAACGCGTG	A		

	1	<u>MLIDYSHSF</u>	<u>FSVPPFLAL</u>	<u>ALAVITRRVL</u>	<u>LSLGIGILVG</u>	<u>VAFLVGGNPV</u>
	51	<u>DGLTHLKMV</u>	<u>VGLAWSGDW</u>	<u>SLGKPKXLVF</u>	<u>LILLGIFTSL</u>	<u>LTYSGSGQAF</u>
	101	<u>ADWAKRHIK</u>	<u>RRGAKMLTAC</u>	<u>LVFVTFIDDY</u>	<u>FHSLAVGAXA</u>	<u>RPVTDKFKVS</u>
40	151	<u>RAKLAYILDS</u>	<u>TAAPMCVLMP</u>	<u>VSSWGASIIA</u>	<u>TLAGLLVTYK</u>	<u>ITEYTPMGTF</u>
	201	<u>VAMSLMNYA</u>	<u>LFALIMVEVV</u>	<u>AWFSFDIGSM</u>	<u>ARFEQAALNE</u>	<u>AHDETAVSDG</u>
	251	<u>SWGVRVALII</u>	<u>PVLALIASTV</u>	<u>SAMIYTGAAQ</u>	<u>SETFSILGAF</u>	<u>ENTDVNTSLV</u>
	301	<u>FGGTCGLAV</u>	<u>VLCTLGTIKI</u>	<u>ADYPKAVWQG</u>	<u>AKSMFGAIAI</u>	<u>LILAWLITSL</u>
	351	<u>VGEMHTGDYL</u>	<u>STLVAGNIHP</u>	<u>GFLXVILELL</u>	<u>AKVMAFATGT</u>	<u>SWGTFGIMLP</u>
	401	<u>IAAAMAVKVD</u>	<u>PSLIIPCMSA</u>	<u>VMAGAVCGDH</u>	<u>CSPISDTTIL</u>	<u>SSTGARCNIH</u>
45	451	<u>DHVTSQLPYA</u>	<u>LTVA AAAAASG</u>	<u>YLALGLTKSA</u>	<u>LLGFGXTGIV</u>	<u>LAVLIFLLKD</u>
	501	KKRANA*				

		10	20	30	40	50	60
50	orf26a.pep	MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILVGVAFLVGGNPVDGLTHLKD	MV				
	orf26-1	MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILVGVAFLVGGNPVDGLTHLKD	MV				
		10	20	30	40	50	60
55	orf26a.pep	VGLAWSGDGWSLGGPKXVLVFLILLGIFTSLTTYSGSNQAFADWAKRHIKNRRGAKMLT	AC				
	orf26-1	VGLAWSGDGWSLGGPKILVFLILLGIFTSLTTYSGSNQAFADWAKRHIKNRRGAKMLT	AC				
		70	80	90	100	110	120
60	orf26a.pep	LVFVTFIDDDYFHS LAVGAXARPVTDKFKVSR	AKLAYILDSTAAPMCVLM	PVSSW	GASII	IA	
	orf26-1	LVFVTFIDDDYFHS LAVGAIA	RPVTDKFKVSR	TAKLAYILDSTAAPMCVLM	PVSSW	GASII	IA
		130	140	150	160	170	180

		130	140	150	160	170	180
5	orf26a.pep	190	200	210	220	230	240
	orf26-1	190	200	210	220	230	240
10	orf26a.pep	250	260	270	280	290	300
	orf26-1	250	260	270	280	290	300
15	orf26a.pep	310	320	330	340	350	360
	orf26-1	310	320	330	340	350	360
20	orf26a.pep	370	380	390	400	410	420
	orf26-1	370	380	390	400	410	420
25	orf26a.pep	430	440	450	460	470	480
	orf26-1	430	440	450	460	470	480
30	orf26a.pep	490	500				
	orf26-1	490	500				

### Homology with a predicted ORF from *N. gonorrhoeae*

- 40 ORF26 shows 94.8% and 99% identity in 97 and 206 aa overlap at the N-terminus and C-terminus, respectively, with a predicted ORF (ORF26ng) from *N. gonorrhoeae*:

	orf26.pep	MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILXXVAFVVGPNPVDGLTHLKDMV	60
	orf26ng	MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILVGVAFVVGPNPVDGLTHLKDMV	60
45	orf26.pep	VGLAWSXDWSLGGPKILVFXILLGIFTSLLTYSGSN	97
	orf26ng	VGLAWADGDWSLGGPKILVFXILLGIFTSLLTYSGSNQAFADWAKRHIKNRCGAKMLTAC	120
50		//	
	orf26.pep	TSLVFGGTCGVFAVVLCTLGTIKTADYPKA	326
	orf26ng	ASTVSAMIYTGAQASETF SILGAFENTDVNTSLVFGGTCGVFAVVLCTFGTIKTADYPKA	326
55	orf26.pep	VWQGA KSMFGAIAILILAWLISTVVGEMHTGDYLS TLVAGNIHPGFLPVILFLLASVMAF	386
	orf26ng	VWQGA KSMFGAIAILILAWLISTVVGEMHTGDYLS TLVAGNIHPGFLPVILFLLASVMAF	386
60	orf26.pep	ATGTSWGTFGIMLP IAAAMAVKVEPALIIPCMSAVMAGAVCGDHCSPI SDTTILSSTGAR	446
	orf26ng	ATGTSWGTFGIMLP IAAAMAVKVEPALIIPCMSAVMAGAVCGDHCSPI SDTTILSSTGAR	446
65	orf26.pep	CNHIDHVTSQLPYALTVA AAAASGYLALGLTKSALLGFGTTGIVLAVLIFLLKDKK	502
	orf26ng	CNHIDHVTSQLPYALTVA AAAASGYLALGLTKSALLGFGTTGIVLAVLIFLLKDKK RADV	506

The complete length ORF26ng nucleotide sequence <SEQ ID 695> is:

```
1 ATGCAGCTGA TTGACTATTC ACATTCATTT TTCTCGGTTG TGCCACCCTT
51 TTTGGCACTG GCACTTGCCG TCATTACCGG CCGCGTACTG CTGTCTTTAG
101 GCATCGGTAT TTTGGTCCGG GTTGCCTTT TGGTCCGGCG CAACCCCGTC
5 GACGGTCTGA CACACCTGAA AGACATGGTC GTCGGCTTGG CTGGGCAGA
201 CGGCGATTGG TCGCTGGGCA AACCAAAAT CTTGGTTTTT CTGATACTTT
251 TGGGCATTTT CACTTCACTG CTGACCTACT CCGGCAGCAA TCAGGCGTTT
301 GCCGACTGGG CAAAACGGCA CATTAAAAAC CGGTGCGGCG CGAAAATGCT
351 GACCGCCTGC CTCGTGTTTC TAACCTTTAT CGACGACTAT TTCCACAGCC
10 401 TCGCCGTCGG TCGGATTGCC CGCCCCGTTA CCGACAAGTT TAAAGTTTCC
451 CGCGCCAAAC TCGCCTACAT CCTCGACTCC ACTGCCTCGC CCATGTGCGT
501 GCTGATGCCC GTTTCAGACT GGGGCGCGTC GATTATCGCC ACGCTTGCCG
551 GATTGCTCGT TACCTACAAA ATTACCGAAT ACACGCCGAT GGGGACGTTT
601 GTCGCCATGA GCCTGATGAA CTATTACGCG CTGTTTGCCC TGATTATGTT
15 651 ATTCGTCGTC GCATGGTTCT CCTTCGACAT CGGCTCGatg gCGCGTTTCG
701 AACAGGCTGC GTTGAACGAA gccaggacg aaaccgccgc tTCAGACgCT
751 ACCAAAGGTC GTGTTTACGC ATTGATTATT CCCGTTTTGG CCTTAATCGC
801 CTCAACGGTT TCCGCCATGA TCTACACCGG CGCGCAGGCA AGCGAAACCT
851 TCAGCATTTT GGGGGCATTT GAAAATACCG ACGTAAACAC TTCGCTGGTA
901 TTCGGCGGCA CTTCGGGCGT GCTTGCCGTC GTCTCTGCA CGTTCGGCAC
20 951 GATTAAACC GCGATTATC CCAAAGCCGT GTGGCAGGT GCGAAATCCA
1001 TGTTCCGGCG AATCGCCATT TTAATCCTCG CCTGGCTCAT CAGTACGGTT
1051 GTCGGCGAAA TGCACACGGG CGACTACCTC TCCACGCTGG TTGCGGGCAA
1101 CATCCATCCC GGCTTCCTGC CCGTCATCCT CTTCCTGCTC GCCAGCGTGA
25 1151 TGGCGTTTGC CACAGGCACA AGCTGGGGGA CGTTCGGCAT TATGCTGCCG
1201 ATTGCCGCCG CCATGGCGGT CAAAGTCGAA CCCGCGCTGA TTAAtcccGTG
1251 TATGTCCGCA GTAATGGCGG GGGCGGTATG CGGCGACCAC TGTTCCGCCA
1301 TCTCCGACAC GACCATCCTG TCGTCCACCG GCGCGCGCTG CAACCACATC
1351 GACCACGTTA CCTCGCAACT GCCTTATGCC CTGACGGTTG CCGCCGCCCG
30 1401 CGCATCGGGC TACCTCGCAT TGGGTCTGAC AAAATCCGCG CTGTTGGGCT
1451 TTGGCACGAC CGGTATTGTA TTGGCGGTGC TGATTTTCT GTTGAAGAT
1501 AAAAAACGCG CCGACGTTTG A
```

This encodes a protein having amino acid sequence <SEQ ID 696>:

```
1 MQLIDYSHSF FSVVPPFLAL ALAVITRRVL LSLGIGILVG VAFVVGPNPV
35 51 DGLTHLKDMV VGLAWADGDW SLGPKILVF LILLGIFTS LTYSGSNQAF
101 ADWAKRHIK RCGAKMLTAC LVFVTFIDY FHSLAVGAI RPVTDKFKVS
151 RAKLAYILDS TASPMCVLMP VSSWGASIIA TLAGLLVYK ITEYTPMGTF
201 VAMSLMNYA LFALIMVFVW AWFSDIGSM ARFEQAALNE AQDETAASDA
251 TKGRVYALII PVLALIASTV SAMIYTGQA SETFSILGAF ENTDVNTSLV
40 301 FGGTCGVLA VLTCTFGTIKT ADYPKAVWQG AKSMFGAIAI LILAWLISTV
351 VGEMHTGDYL STLVAGNIHP GFLLPVILFL ASVMAFATGT SWGTFGIMLP
401 IAAAMAVKVE PALIIPCMSA VMAGAVCGDH CSPISDTTIL SSTGARCNIH
451 DHVTSQLPYA LTVAAAAASG YLALGLTRSA LLFGFTTGIV LAVLIFLLKD
501 KKRADV*
```

ORF26ng and ORF26-1 show 98.4% identity in 505 aa overlap:

```
10 20 30 40 50 60
orf26-1.pep MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILVGVAFVVGPNPVDGLTHLKDMV
50 orf26ng MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILVGVAFVVGPNPVDGLTHLKDMV
10 20 30 40 50 60
70 80 90 100 110 120
orf26-1.pep VGLAWSGDGWSLGLPKILVFLILLGIFTSLLTYSGSNQAFADWAKRHIKNRRGAKMLTAC
55 orf26ng VGLAWADGWSLGLPKILVFLILLGIFTSLLTYSGSNQAFADWAKRHIKNRCGAKMLTAC
70 80 90 100 110 120
130 140 150 160 170 180
orf26-1.pep LVFVTFIDYFHSLAVGAIARPVTDKFKVSRTKLAYILDSTAAPMCVLPVSSWGASIIA
60 orf26ng LVFVTFIDYFHSLAVGAIARPVTDKFKVSRAKLAYILDSTASPMCVLPVSSWGASIIA
130 140 150 160 170 180
190 200 210 220 230 240
```



5	orf26-1.pep	TLAGLLVITYKITEYTPMGTFFVAMSLMNYALFALIMVFVAVWFSFDIGSMARFEQAALNE
	orf26ng	TLAGLLVITYKITEYTPMGTFFVAMSLMNYALFALIMVFVAVWFSFDIGSMARFEQAALNE
10	orf26-1.pep	AHDETAVSDATKGRVYALIIPVLALIASTVSAMIYTGAQASETFSILGAFENTDVNTSLV
	orf26ng	AQDETAASDATKGRVYALIIPVLALIASTVSAMIYTGAQASETFSILGAFENTDVNTSLV
15	orf26-1.pep	FGGTCGVLAVVLTCTGTIKTADYPKAVWQGAQSMFGAIAILILAWLISTVVGEMHTGDYL
	orf26ng	FGGTCGVLAVVLTCTGTIKTADYPKAVWQGAQSMFGAIAILILAWLISTVVGEMHTGDYL
20	orf26-1.pep	STLVAGNIHPGFLPVILFLLASVMAFATGTSWGTFGIMLPAAAMAVKVEPALIIPCMSA
	orf26ng	STLVAGNIHPGFLPVILFLLASVMAFATGTSWGTFGIMLPAAAMAVKVEPALIIPCMSA
25	orf26-1.pep	VMAGAVCGDHCSPISTDTTILSSTGARCNIHDVTSQLPYALTVA AAAASGYLALGLTKSA
	orf26ng	VMAGAVCGDHCSPISTDTTILSSTGARCNIHDVTSQLPYALTVA AAAASGYLALGLTKSA
30	orf26-1.pep	LLGFGTTGIVLAVLIFLLKDKKRANAX
	orf26ng	LLGFGTTGIVLAVLIFLLKDKKRADVX

35 In addition, ORF26 ng shows significant homology to a hypothetical *H. influenzae* protein:

40	sp P44263 YF86_HAEIN	HYPOTHETICAL PROTEIN HI1586	>gi 1074850 pir  C64037
	hypothetical	protein HI1586 - Haemophilus influenzae (strain Rd KW20)	>gi 1574427 (U32832) H. influenzae predicted coding region HI1586 [Haemophilus influenzae] Length = 519
45	Score =	538 bits (1370), Expect = e-152	
	Identities =	280/507 (55%), Positives = 346/507 (68%), Gaps = 7/507 (1%)	
50	Query: 1	MQLIDYSHSFFSVVPPFLALALAVITRRXXXXXXXXXXXXXAFVVGPNVDGLTHLKDMV	60
	Sbjct: 14	MELIDFSSSVWSIVPALLAIIAIAITRRVLVSLSAGIIGSMLSDWQIGSAFNVLVKNV	73
55	Query: 61	VGLAWADGDWSLGPVILVFLILLGIFTSLLTSGSNQAFADWAKRHIKNRCGAKMLTAC	120
	Sbjct: 74	VSLVYADGEIN-SNMNIVLFLLLGLVLTALLTVSGSNRAFAEWAQSRIGRRGAKLLAAS	132
60	Query: 121	LVFVTFIDDFHSLAVGAIARPVTDKFKVSRKALAYILDSTASPMCVLMPVSSWGASIIA	180
	Sbjct: 133	LVFVTFIDDFHSLAVGAIARPVTDKFKVSRKALAYILDSTA+PMCV+MPVSSWGA II	192
65	Query: 181	TLAGLLVITYKITEYTPMGTFFVAMSLMNYALFALIMVFVAVWFSFDIGSMARFEQAALNE	240
	Sbjct: 193	LIGGLLATYSITEYTPIGAFVAMSSMNFYAIFSIIMVFFVAYFSFDIASMVRHEKLALKN	252
70	Query: 241	AQDETAASDATKGRVYALIIPVLALIASTVSAMIYTGAQA---SETFSILGAFENTDVN	296
	Sbjct: 253	TGQLEEETGKQVRNLILPILVLIATVSMMIYTGAELAADGKVFVSLGTFFENTVVG	312
75	Query: 297	TSLVFGGTCGVL--AVVLTCTGTIKTADYPKAVWQGAQSMFGXXXXXXXXXXSTVVGEM	354
	Sbjct: 313	TSLVGGC++ +++ + +Y++ G KSM G + +VG+M	372
80	Query: 355	HTGDYLSLTVAGNIHPGFLPVILFLLASVMAFATGTSWGTFGIMLPAAAMAVKVEPALI	414
		TG YLS+LV+GNI FLPVILF+L + MAF+TGTSWGTFGIMLPAAAMA P L+	

Sbjct: 373 QTGKYLSSLVSGNIPMQFLPVILFVLGAAMAFSTGTSWGTGIMLPFAAAMAANAPELL 432  
 Query: 415 IPCMSAVMAGAVCGDHCSPISDITLSTGARC�HIDHVTSQXXXXXXXXXXXXXXXXXXXX 474  
 +PC+SAVMAGAVCGDHCSP+SDITLSTGA+CNHIDHVT+Q  
 Sbjct: 433 LPCLSAVMAGAVCGDHCSPVSDITLSTGAKCNHIDHVTTQLPYAATVATATSIGYIVV 492  
 Query: 475 XXXKSALLGFGTTGIVLAVLIFLLKDK 501  
 S L G F T + L V + I F + K +  
 Sbjct: 493 GFTYSGLAGFAATAVSLIVIIFAVKKR 519

Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 83

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 697>:

15 1 ..AAGCAATGGT ATGCCGACGN .AGTATCAAG ACGGAAATGG TTATGGTCAA  
 51 CGATGAGCCT GCCAAATTC TGACTTGGGA TGAAAGCGGC CGATTACTCT  
 101 CGGAAGTGT TATCCGCCAC CATCAACGCA ACGGGGTGGT TTTGGAGTGG  
 151 TATGAAGATG GTTCTAAAAA GAGCGAAGT. GTTTATCAGG ATGACAAGTT  
 201 GGTCAAGAAA ACCCAGTGGG ATAAGGATGG TTATTTAATC GAACCTGA

20 This corresponds to the amino acid sequence <SEQ ID 698; ORF27>:

1 ..KQWYADXSIK TEMVMVNDEP AKILTWDESG RLLSELSIRH HQRNGVVLEW  
 51 YEDGSKKSEX VYQDDKLVRK TQWDKDGyli EP\*

Further work revealed the complete nucleotide sequence <SEQ ID 699>:

25 1 ATGAAAAAAT TATCTCGGAT TGTATTTTCA ACTGTCCTGT TGGGTTTTTC  
 51 GGCCGCTTTG CCGGCGCAGA CCTATTCTGT TTATTTTAAAT CAGAACGGAA  
 101 AGCTGACGGC GACGATGTCT TCTGCCGCTT ATATCAGGCA ATATAGTGTG  
 151 GTGGCGGGTA TTGCGCACGC GCAGGATTTT TATTATCCGT CGATGAAGAA  
 201 ATATTCTGAA CCTTATATCG TTGCTTCAAC GCAAATCAAA TCTTTGTGTC  
 251 CTACCCCTGCA AAACGGTATG TTGATTTTGT GGCATTTTAA TGGTCAGAAA  
 301 AAAATGCGCG GGGGCTTCAG CAAGGGTAAG CCGGACGGGG AGTGGGTCAA  
 351 CTGGTATCCG AACGGTAAAA AATCTGCCGT TATGCCTTAT AAAAATGGCT  
 401 TGAGTGAGGG TACGGGATAC CGTATTACC GTAACGGCGG CAAGGAAAGC  
 451 GAAATCCAGT TTAAGCAAAA TAAGGCAAAC GGCGTATGGA AGCAATGGTA  
 501 TGCCGACGGC AGTATCAAGA CGGAAATGGT TATGGTCAAC GATGAGCCTG  
 35 551 CCAAAATTCT GACTTGGGAT GAAAGCGGCC GATTACTCTC GGAAGTGTCT  
 601 ATCCGCCACC ATCAACGCAA CGGGGTGGTT TTGGAGTGGT ATGAAGATGG  
 651 TTCTAAAAAG AGCGAAGCTG TTTATCAGGA TGACAAGTTG GTCAGGAAAA  
 701 CCCAGTGGGA TAAGGATGGT TATTTAATCG AACCTGA

This corresponds to the amino acid sequence <SEQ ID 700; ORF27-1>:

40 1 MKKLSRIVFS TVLLGFSAAL PAQTYSVYFN QNGKLTATMS SAAYIROYSV  
 51 VAGIAHAQDF YYPSMKKYSE PYIVASTQIK SFVPTLQNGM LILWHFNGQK  
 101 KMAGGFSKKG PDGEWVNWYP NGKKSAMVYP KNGLSEGTY RYYRNGGKES  
 151 EIQFKQNKAN GVWKQWYADG SIKTEMVMVN DEPAKILTWD ESGRLLSELS  
 201 IRHHQRNGVV LEWYEDGSKK SEAVYQDDKL VRKTQWDKDG YLIEP\*

45 Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF27 shows 91.5% identity over a 82aa overlap with an ORF (ORF27a) from strain A of *N.meningitidis*:

50 orf27.pep 10 20 30  
 KQWYADXSIKTEMVMVNDEPAKILTWDESG

or27a LSEGTGXRYRNGGKESEIQFKQNKANGVWKQWYADGNIKTEMVMVNDEPAKILTWDSESG  
140 150 160 170 180 190

5 orf27.ppe RLLSELSIRHHQRNGVVLEWYEDGSKKSEXVYQDDKLVRKTQWDKDGYLIEPX  
40 50 60 70 80  
or27a RLLSELSIHHRNGVVLEWYEDGSKKXEAUVYQDDKLVRKTQWDKDGYLIEPX  
200 210 220 230 240

10 The complete length ORF27a nucleotide sequence <SEQ ID 701> is:

1 ATGAAAAAAT TATCTCGGAT TGTATTTTCA ACTGTCCTGT TGGGTTTTTC  
51 GGCCGCTTTG CCGGCGCAGA NCTATTCTGT TTATTTTAAT CAGAACGGGA  
101 AACTGACGGC GACGNTGTCT TCTGCCGNT ATATCAGGCA ATATAGTGTG  
151 GCGGAGGGTA TTGCGCACGC GCAGGANTTT TANTATCCGT CGATGAAGAA  
15 201 ATATTCCGAA CCTTATATCG TTGCTTCAAC GCAAATCAAA TCTTTTGTGC  
251 CTACCCTGCA AAACGGTATG TTGATTTTGT GGCATTTTAA NGGTCAGAAA  
301 AAAATGGCNG GGGGCTTCAG CAAGGGTAAG CCGGACGGGG AGTGGGTCAA  
351 CTGGTATCCG AACGGTAAAA AATCTGCCGT TATGCCTTAT AAAAATGGTT  
401 TGAGTGAAGG TACGGGGTNN CGCTATTACC GTACGGCGG CAAGGAAAGC  
20 451 GAAATCCAGT TTAACAGAA TAAGGCAAAC GCGGTATGGA AGCAATGGTA  
501 TGCCGACGGC AATATCAAAA CGGAAATGGT TATGGTCAAT GATGAGCCTG  
551 CCAAATTCT GACATGGGAT GAAAGCGGTC GATTACTCTC GGAAGTCTCT  
601 ATCCATCATC ATNAACGTAA TGGAGTAGTC TTAGAGTGGT ATGAAGATGG  
151 TTCTAAAAAG ANTGAAGCTG TTTATCAGGA TGATAAGTTG GTCAGGAAAA  
25 701 CCAGTGGGA TAANGATGGT TATTTAATCG AACCTGA

This encodes a protein having amino acid sequence <SEQ ID 702>:

1 MKKLSRIVFS TVLLGFSAAL PAQXYSVYFN QNGKLTATXS SAAYIRQYSV  
51 AEGIAHAQXF XYPMSMKYSE PYIVASTQIK SFVPTLQNGM LILWHFXGQK  
101 KMAGGFSKSK PDGEVWNWYP NGKKSAMVMPY KNLSEGTGX RYYRNGGKES  
30 151 EIQFKQNKAN GVWKQWYADG NIKTEMVMVN DEPAKILTWD ESGRLLSELS  
201 IHHHRNGVV LEWYEDGSKK XEAUVYQDDKL VRKTQWDKDG YLIEP\*

ORF27a and ORF27-1 show 94.7% identity in 245 aa overlap:

35 or27a.ppe 10 20 30 40 50 60  
MKKLSRIVFSTVLLGFSAALPAQXYSVYFNQNGKLTATXSAAAYIRQYSVAEGIAHAQXF  
or27-1 10 20 30 40 50 60  
MKKLSRIVFSTVLLGFSAALPAQYSVYFNQNGKLTATMSSAAYIRQYSVVAGIAHAQDF

40 or27a.ppe 70 80 90 100 110 120  
XYPMSMKYSEPYIVASTQIKSFVPTLQNGMLILWHFXGQKKMAGGFSKSKPDGEVWNWYP  
or27-1 70 80 90 100 110 120  
YYPMSMKYSEPYIVASTQIKSFVPTLQNGMLILWHFNGQKKMAGGFSKSKPDGEVWNWYP

45 or27a.ppe 130 140 150 160 170 180  
NGKKSAMVMPYKNLSEGTGXRYRNGGKESEIQFKQNKANGVWKQWYADGNIKTEMVMVN  
or27-1 130 140 150 160 170 180  
NGKKSAMVMPYKNLSEGTGYRYRNGGKESEIQFKQNKANGVWKQWYADGSIKTEMVMVN

50 or27a.ppe 190 200 210 220 230 240  
DEPAKILTWDSEGRLLSELSIHHRNGVVLEWYEDGSKKXEAUVYQDDKLVRKTQWDKDG  
or27-1 190 200 210 220 230 240  
DEPAKILTWDSEGRLLSELSIRHHQRNGVVLEWYEDGSKKXEAUVYQDDKLVRKTQWDKDG

55 or27a.ppe YLIEPX  
or27-1 YLIEPX

60

Homology with a predicted ORF from *N.gonorrhoeae*

ORF27 shows 96.3% identity over 82 aa overlap with a predicted ORF (ORF27ng) from *N.gonorrhoeae*:

```

5      orf27.pep                                KQWYADXSIKTEMVMVNDEPAKILTWDESG    30
      orf27ng      LSEGTYGYRYYRNGGKESEIQFKQNKANGVWKQWYADGSIKTEMVMVNDEPAKILTWDESG    193
      orf27.pep      RLLSELSIRHHQRNGVVLEWYEDGSKKSEAVYQDDKLVRKTQWDKDGYLIEP    82
10     orf27ng      RLLSELSIRHHKRNGVVLEWYEDGSKKSEAVYQDDKLVRKTQWDKDGYLIEP    245

```

The complete length ORF27ng nucleotide sequence <SEQ ID 703> is:

```

1  ATGAAGAAAT TATCTCGGAT TGTATTTTCA ATCGTACTGT TGGGTTTTTC
51 GGCCGCTTTG CCGGCGCAGA CCTATTCTGT TTATTTTAAT CAGAACGGGA
101 AACTGACGGC GACGATGTCT TCTGCCGCTT ATATCAGGCA ATATAGTGTG
151 GCGGCGGGTA TCGCACACGC GCAGGATTTT TATTATCCGT CGATGAAGAA
201 ATATTCCGAA CCTATATATCG TTGCTTCAAC GCAAATCAAA TCCTTTGTGC
251 CTACCCGTCA AAACGGTATG TTGATTTTGT GGCATTTTAA TGGTCAGAAA
301 AAAATGGCGG GGGGCTTCAG CAAGGGTAAG CCGACGGGG AATGGGTCAA
351 CTGGTATCCG AACGGTAAAA AATCTGCGGT TATGCCTTAT AAAAATGGCT
20 401 TGAGTGAGGG TACGGGATAC CGTTATTACC GTAACGGCGG CAAGGAAAGC
451 GAAATCCAGT TTAAGCAAAA TAAGGCGAAC GCGGTATGGA AGCAATGGTA
501 TGCCGATGGA AGTATCAAGA CGGAAATGGT TATGGTCAAC GATGAGCCTG
551 CCAAAATTCT GACTTGGGAT GAAAGCGGCC GATTACTTTC GGAAGTGTCT
601 ATCCGCCACC ATAAACGCAA CGGGGTGGTT TTGGAGTGGT ATGAAGATGG
25 651 TTCTAAAAAG AGCGAGGCTG TTTATCAGGA TGACAAGTTG GTCAGGAAAA
701 CCAATGGGA TAAGGATGGT TATTTAATCG AACCTGA

```

This encodes a protein having amino acid sequence <SEQ ID 704>:

```

1  MKKLSRIVES IVLLGFSAAL PAQTYSVYFN QNGKLTATMS SAAYIRQYSV
51 AAGIAHAQDF YYPSMKKYSE PYIVASTQIK SFVPTLQNGM LILWHFNGQK
30 101 KMAGGFSK GK PDGEWVNWYP NGKKSAMVMPY KNGLSEGTGY RYYRNGGKES
151 EIQFKQNKAN GVWKQWYADG SIKTEMVMVN DEPAKILTWD ESRLLSELS
201 IRHHRNGVV LEWYEDGSKK SEAVYQDDKL VRKTQWDKDG YLIEP*

```

ORF27ng and ORF27-1 show 98.8% identity in 245 aa overlap:

```

35      orf27-1.pep      10      20      30      40      50      60
      orf27ng      MKKLSRIVFSTVLLGFSAALPAQTYSVYFNQNGKLTATMSSAAYIRQYSVVAAGIAHAQDF
      orf27ng      MKKLSRIVFSIVLLGFSAALPAQTYSVYFNQNGKLTATMSSAAYIRQYSVAAGIAHAQDF
      orf27-1.pep      70      80      90      100     110     120
      orf27ng      YYPSMKKYSEPYIVASTQIKSFVPTLQNGMLILWHFNGQKKMAGGFSKGPGEWVNWYP
      orf27ng      YYPSMKKYSEPYIVASTQIKSFVPTLQNGMLILWHFNGQKKMAGGFSKGPGEWVNWYP
45      orf27-1.pep      130     140     150     160     170     180
      orf27ng      NGKKSAMVMPYKNGLSEGTGYRYYRNGGKESEIQFKQNKANGVWKQWYADGSIKTEMVMVN
      orf27ng      NGKKSAMVMPYKNGLSEGTGYRYYRNGGKESEIQFKQNKANGVWKQWYADGSIKTEMVMVN
50      orf27-1.pep      190     200     210     220     230     240
      orf27ng      DEPAKILTWDESGRLLSELSIRHHQRNGVVLEWYEDGSKKSEAVYQDDKLVRKTQWDKDG
      orf27ng      DEPAKILTWDESGRLLSELSIRHHRNGVVLEWYEDGSKKSEAVYQDDKLVRKTQWDKDG
55      orf27-1.pep      190     200     210     220     230     240
      orf27ng      DEPAKILTWDESGRLLSELSIRHHRNGVVLEWYEDGSKKSEAVYQDDKLVRKTQWDKDG
60      orf27-1.pep      YLIEPX
      orf27ng      YLIEPX

```

orf27ng

YLIEPX

Based on this analysis, including the putative leader sequence in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF27-1 (24.5kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 17A shows the results of affinity purification of the GST-fusion protein, and Figure 17B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA, which gave a positive result, confirming that ORF27-1 is a surface-exposed protein and a useful immunogen.

#### Example 84

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 705>:

```

1  ATGAAATTTA CCAAGCACCC CGTCTGGGCA ATGGCGTTCC GCCCATTTTA
51  TTCGCTGGCG GCTCTGTACG GCGCATTGTC CGTATTGCTG TGGGGTTTCG
101 GCTACACGGG AACGCACAGG CTGTCCGGTT TCTATTGGCA CGCGCATGAG
151 ATGATTGGGG GTTATGCCGG ACTGGTCGTC ATCGCCTTCC TGCTGACCGC
201 CGTCGCCACT TGGACGGGGC AGCCGCCCAC GCGGGGCGGC GTaTCTGGTC
251 GGCTTGACTA TCTTTTGGCT GGCTGCGCGG ATTGCCGCCT TTATCCCGGG
301 TTGGGGTGCG TCGGCAAGCG GCATACTCGG TACGCTGTTT TTCTGGTACG
351 GCGCGGTGTG CATGGCTTTG CCCGTTATCC GTTCGCAGAA TCAACGCAAC
401 TATGTTgCCG TGTTGCGGCT GTTCGTCTTG GCGGCGACGC ATGCGGCGTT
451 CCACGTCCAG CTGCACAACG GCAACCTAGG CGGACTCTTG AGCGGATTGC
501 AGTCGGGCTT GGTGATG

```

This corresponds to the amino acid sequence <SEQ ID 706; ORF47>:

```

1  MKFTKHPVWA MAFRPFYSLA ALYGALSVLL WFGYTGTHX LSGFYWHAHE
51  MIWGYAGLVV IAFLLTAVAT WTGQPPTRGV VLVGLTIFWL AARIAAFIPG
101 WGASASGILG TLFFWYGAVC MALPVIRSQN QRNYVAVFAL FVLGGTHAAF
151 HVQLHNGNLG GLLSGLQSGL VM

```

Further work revealed the complete nucleotide sequence <SEQ ID 707>:

```

1  ATGAAATTTA CCAAGCACCC CGTCTGGGCA ATGGCGTTCC GCCCATTTTA
51  TTCGCTGGCG GCTCTGTACG GCGCATTGTC CGTATTGCTG TGGGGTTTCG
101 GCTACACGGG AACGCACGAG CTGTCCGGTT TCTATTGGCA CGCGCATGAG
151 ATGATTGGGG GTTATGCCGG ACTGGTCGTC ATCGCCTTCC TGCTGACCGC
201 CGTCGCCACT TGGACGGGGC AGCCGCCCAC GCGGGGCGGC GTTCTGGTCG
251 GCTTGACTAT CTTTGGGCTG GCTGCGCGGA TTGCCGCCTT TATCCCGGGT
301 TGGGGTGCGT CGGCAAGCGG CATACTCGGT ACGCTGTTTT TCTGGTACGG
351 GCGGGTGTCG ATGGCTTTGC CCGTTATCCG TTCGCAGAAAT CAACGCAACT
401 ATGTTGCCGT GTTCGCGCTG TTCGTCTTGG GCGGCGACGA TGCGGGCTTC
451 CACGTCCAGC TGCAACAACG CAACCTAGGC GGACTCTTGA GCGGATTGCA
501 GTCGGGCTTG GTGATGGTGT CCGGTTTTAT CCGTCTGATT GGTACGCGGA
551 TTATTTTCGT TTTTACGTCC AAACGCTTGA ATGTGCCGCA GATTCCAGT
601 CCGAAATGGG TGGCGCAGGC TTCGCTGTGG CTGCCCATGC TGAATGCCAT
651 GCTGATGGCG CACGGTGTGT TGGCTTGGCT GTCTGCCGTT TTTGCCTTTG
701 CGGCAGGTGT GATTTTTTACC GTGCAGGTGT ACCGCTGGTG GTATAAACC
751 GTGTTGAAAG AGCCGATGCT GTGATTCTG TTTGCCGGCT ATCTGTTTAC
801 CGGATTGGGG CTGATTGCGG TCGGCGCGTC TTATTTCAAA CCCGCTTTCC
851 TCAATCTGGG TGTGCATCTG ATCGGGGTCG GCGGTATCGG CGTGCTGACT
901 TTGGGCATGA TGGCGCGTAC CGCGCTTGGT CATACGGGCA ATCCGATTTA

```

5  
 951 TCCGCCGCCC AAAGCCGTTT CCGTTGCGTT TTGGCTGATG ATGGCGGCAA  
 1001 CCGCCGTCCG TATGGTTGCC GTATTTTCTT CCGGCACTGC CTACACGCAC  
 1051 AGCATCCGCA CCTCTTCGGT TTTGTTTGCA CTCGCGCTT TGGTGTATGC  
 1101 GTGGAAGTAT ATTCCTTGGC TGATTCGTCC GCGTTCGGAC GGCAGGCCCC  
 1151 GTTGA

This corresponds to the amino acid sequence <SEQ ID 708; ORF47-1>:

10  
 1 MKFTKHPVWA MAFRPFYSLA ALYGALSVLL WFGYTGTHE LSGFYWHAHE  
 51 MIWGYAGLVV IAFLLTAVAT WTGQPPTRGV VLVGLTIFWL AARIAAFIPG  
 101 WGASASGILG TLFFWYGAVC MALPVIRSON QRNYVAVFAL FVLGGTHAAF  
 151 HVQLHNGNLG GLLSGLQSGV VMVSGFIGLI GTRIIISFFTS KRLNVPQIPS  
 201 PKWVAQASLW LPMLTAMLMA HGVLAWSAV FAFAGVIFT VQVYRWYKYP  
 251 VLKEPMLWIL FAGYLTGLG LIAVGASYFK PAFLNLGVHL IGVGGIGVLT  
 301 LGMMARTALG HTGNPIYPPP KAVPVAFWLM MAATAVRMVA VFSSGTAYTH  
 351 SIRTSSVLFA LALLVYAWKY IPWLIRPRSD GRPG\*

15 Computer analysis of this amino acid sequence predicts a leader peptide and also gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF47 shows 99.4% identity over a 172aa overlap with an ORF (ORF47a) from strain A of *N. meningitidis*:

20  
 orf47.pep MKFTKHPVWMAFRPFYSLAALYGALSVLLWFGYTGTHXLSGFYWHAHEMIWGYAGLVV  
 orf47a MKFTKHPVWMAFRPFYSLAALYGALSVLLWFGYTGTHELSGFYWHAHEMIWGYAGLVV  
 25  
 orf47.pep IAFLLTAVATWTGQPPTRGVVLVGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC  
 orf47a IAFLLTAVATWTGQPPTRGVVLVGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC  
 30  
 orf47.pep MALPVIRSONQRNYVAVFALFVLGGTHAAFHVQLHNGNLGGLSGLQSGLVLM  
 orf47a MALPVIRSONQRNYVAVFALFVLGGTHAAFHVQLHNGNLGGLSGLQSGLVLMVSGFIGLI  
 35  
 orf47a GTRIIISFFTSKRLNVPQIPSPKWVAQASLWLPMLTAMLMAHGVMPWLSAFAFAAGVIFT

40 The complete length ORF47a nucleotide sequence <SEQ ID 709> is:

1 ATGAAATTTA CCAAGCACCC CGTTTGGGCA ATGGCGTTCC GCCCGTTTTA  
 51 TTCCTGGCG GCTCTGTACG GCGCATTTGC CGTATTGCTG TGGGGTTTCG  
 101 GCTACACGGG AACGCACGAG CTGTCCGGTT TCTATTGGCA CGCGCATGAG  
 151 ATGATTTGGG GTTATGCCGG ACTGGTCGTC ATCGCCTTCC TGCTGACCGC  
 45 201 CGTCGCCACT TGGACGGGGC AGCCGCCAC GCGGGCGGC GTTCTGGTCG  
 251 GCTTGAATAT CTTTGGCTG GCTGCGCGGA TTGCGCCTT TATCCCGGT  
 301 TGGGGTGCGT CGGCAAGCGG CATACTCGGT ACGCTGTTT TCTGGTACGG  
 351 CGCGGTGTGC ATGGCTTTGC CCGTTATCCG TTCGAGAAT CAACGCAATT  
 401 ATGTTGCCGT GTTCGCGCTG TTCGTCTTGG GCGGTACGCA CGCGCGTTC  
 50 451 CACGTCCAGC TGCACAACGG CAACCTAGGC GGACTCTTGA GCGATTGCA  
 501 GTCGGGCTTG GTGATGGTGT CGGGTTTTAT CGGTCTGATT GGTACGCGGA  
 551 TTATTTCTGT TTTTACGTCC AAACGGTTGA ATGTGCCGCA GATTCACAGT  
 601 CCGAAATGGG TGGCGCAGGC TTCGTGTGG CTGCCATGC TGACCGCAT  
 651 CGTGATGGCG CACGCGTGA TGCCTTGGCT GTCGCGGCT TTCGCGTTG  
 701 CGGCAGGTGT GATTTTACC GTGCAGGTGT ACCGCTGGTG GTATAAGCCT  
 751 GTGTTGAAAG AGCCGATGCT GTGGATTCTG TTTGCCGGCT ATCTGTTTAC  
 801 CGGATTGGGG CTGATTGCGG TCGGCGCGTC TTATTTCAA CCCGCTTTC  
 851 TCAATCTGGG TGTGCATCTG ATCGGGGTCG GCGGTATCGG CGTGCTGACT

5  
901 TTGGGCATGA TGGCGCGTAC CGCGCTCGGT CATACGGGCA ATCCGATTTA  
951 TCCGCCGCC AAAGCCGTT CCGTTGCGTT TTGGCTGATG ATGGCGGCAA  
1001 CCGCCGTCCG TATGGTTGCC GTATTTTCTT CCGGCACTGC CTACACGCAC  
1051 AGCATACGCA CCTCTTCGGT TTTGTTTGCA CTCGCGCTTT TGGTGTATGC  
1101 GTGGAAGTAT ATTCCTTGCC TGATTCGTCC GCGTTCGGAC GGCAGGCCCG  
1151 GTTGA

This encodes a protein having amino acid sequence <SEQ ID 710>:

10  
15  
1 MKFTKHPVWA MAFRPFYSLA ALYGALSULL WGFGYTGTHE LSGFYWHAHE  
51 MIWGYAGLVV IAFLLTAVAT WTGQPPTRG VLVGLTIFWL AARIAAFIPG  
101 WGASASGILG TLFFWYGAVC MALPVIRSON QRNYVAVFAL FVLGGTHAAF  
151 HVQLHNGNLG GLLSGLQSG LVMVSGFIGLI GTRIISFFTS KRLNVPQIPS  
201 PKWVAQASLW LPMLTAMLMA HGVMPWLSAA FAFAGVIFT VQVYRWYKYP  
251 VLKEPMLWIL FAGYLTGLG LIAVGASYFK PAFLNLGVHL IGVGGIGVLT  
301 LGMMARTALG HTGNPIYPPP KAVPVAFWLM MAATAVRMVA VFSSGTAYTH  
351 SIRTSSVLFA LALLVYAWKY IPWLIRPRSD GRPG\*

ORF47a and ORF47-1 show 99.2% identity in 384 aa overlap:

20  
25  
30  
35  
40  
45  
50  
55  
orf47a.pep 10 20 30 40 50 60  
MKFTKHPVWMAFRPFYSLAALYGALSULLWGFGYTGTHELSGFYWHAHEMIWGYAGLVV  
|||||  
orf47-1 10 20 30 40 50 60  
MKFTKHPVWMAFRPFYSLAALYGALSULLWGFGYTGTHELSGFYWHAHEMIWGYAGLVV  
70 80 90 100 110 120  
orf47a.pep IAFLLTAVATWTGQPPTRGVLVGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC  
|||||  
orf47-1 IAFLLTAVATWTGQPPTRGVLVGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC  
70 80 90 100 110 120  
orf47a.pep 130 140 150 160 170 180  
MALPVIRSONQRNYVAVFALFVLGGTHAAAFHVQLHNGNLGGLLSGLQSGLVMSVSGFIGLI  
|||||  
orf47-1 MALPVIRSONQRNYVAVFALFVLGGTHAAAFHVQLHNGNLGGLLSGLQSGLVMSVSGFIGLI  
130 140 150 160 170 180  
orf47a.pep 190 200 210 220 230 240  
GTRIISFFTSKRLNVPQIPSPKWVAQASLWLPMLTAMLMAHGVMPWLSAAFAFAAGVIFT  
|||||  
orf47-1 GTRIISFFTSKRLNVPQIPSPKWVAQASLWLPMLTAMLMAHGVLAWSAVFAFAAGVIFT  
190 200 210 220 230 240  
orf47a.pep 250 260 270 280 290 300  
VQVYRWYKYPVLKEPMLWILFAGYLTGLGLIAVGASYFKPAFLNLGVHLIGVGGIGVLT  
|||||  
orf47-1 VQVYRWYKYPVLKEPMLWILFAGYLTGLGLIAVGASYFKPAFLNLGVHLIGVGGIGVLT  
250 260 270 280 290 300  
orf47a.pep 310 320 330 340 350 360  
LGMMARTALGHTGNPIYPPP KAVPVAFWLMMAATAVRMVAVFSSGTAYTHSIRTSSVLFA  
|||||  
orf47-1 LGMMARTALGHTGNPIYPPP KAVPVAFWLMMAATAVRMVAVFSSGTAYTHSIRTSSVLFA  
310 320 330 340 350 360  
orf47a.pep 370 380  
LALLVYAWKYIPWLIRPRSDGRPGX  
|||||  
orf47-1 LALLVYAWKYIPWLIRPRSDGRPGX  
370 380

Homology with a predicted ORF from *N.gonorrhoeae*

60 ORF47 shows 97.1% identity over 172 aa overlap with a predicted ORF (ORF47ng) from  
*N.gonorrhoeae*:

ORF47 MKFTKHPVWMAFRPFYSLAALYGALSULLWGFGYTGTHELSGFYWHAHEMIWGYAGLVV 60

ORF47ng  
 5 ORF47  
 ORF47ng  
 ORF47  
 10 ORF47ng

```

|||||
MKFTKHPVWMAFRPFYSLAALYGALSVLLWGFYGTGTHELSGFYWHAHEMIWGYAGLVV 60
IAFLLTAVATWTGQPPTRGGLVGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC 120
|||||:|||||
IAFLLTAVATWTGQPPTRGGLVGLTAFWLAARIAAFIPGWGAAASGILGTLFFWYGAVC 120
MALPVIRSQNRNYVAVFALFVLGGTHAAFHVQLHNGNLGGLLSGLQSGLVM 172
|||||:|||||
MALPVIRSQNRNYVAVFAIFVLGGTHAAFHVQLHNGNLGGLLSGLQSGLVMVWGFGLI 180

```

The ORF47ng nucleotide sequence <SEQ ID 711> is predicted to encode a protein comprising amino acid sequence <SEQ ID 712>:

15

```

1 MKFTKHPVWA MAFRPFYSLA ALYGALSVLL WGFYGTGTHE LSGFYWHAHE
51 MIWGYAGLVV IAFLLTAVAT WTGQPPTRGGLVGLTAFWL AARIAAFIPG
101 WGAASGILG TLFFWYGAVC MALPVIRSQNRNYVAVFAI FVLGGTHAAF
151 HVQLHNGNLG GLLSGLQSGL VMVWGFGLI GMKIIISFSTS KRLKLPQIPS
201 PKWVAHASLW LPMLNAILMA HRVMPWLSAA FPFAAGVIFT VQVYAGGITP
251 IEETSCGSVA GICYRLGNSS G

```

20 The predicted leader peptide and transmembrane domains are identical (except for an Ile/Ala substitution at residue 87 and an Leu/Ile substitution at position 140) to sequences in the meningococcal protein (see also *Pseudomonas stutzeri* orf396, accession number e246540):

25

```

TM segments in ORF47ng
INTEGRAL Likelihood = -5.63 Transmembrane 52 - 68
INTEGRAL Likelihood = -3.88 Transmembrane 169 - 185
INTEGRAL Likelihood = -3.08 Transmembrane 82 - 98
INTEGRAL Likelihood = -1.91 Transmembrane 134 - 150
INTEGRAL Likelihood = -1.44 Transmembrane 107 - 123
INTEGRAL Likelihood = -1.38 Transmembrane 227 - 243

```

Further work revealed the complete gonococcal DNA sequence <SEQ ID 713>:

30

```

1 ATGAAATTTA CCAACATCC CGTCTGGGCA ATGGCGTTCC GCCCGTTTTA
51 TTCCTGGCG GCACTGTACG GCGCATTGTC CGTATTGCTG TGGGGTTTCG
101 GCTACACGGG AACGCACGAG CTGTCCGTT TCTATTGGCA CGCGCATGAG
151 ATGATTTGGG GTTATGCCGG TCTCGTCGTC ATCGCCTTCC TGCTGACCCG
201 CGTCGCCACT TGGACGGGAC AGCCGCCAC GAGGGCGGC GTTCTGGTCG
35 GCTTGACCGC CTTTTGGCTG GCTGCGCGGA TTGCCGCCTT TATCCGGGGT
301 TGGGGTGCGG CGGCAAGCGG CATACTCGGT ACGTGTGTTT TCTGGTACGG
351 CGCGGTGTGC ATGGCTTTGC CCGTTATCCG TtcgCAAAAC CGGCGCAACT
401 ATGtcgCCGT ATTcGCAATA TTTGTGCTGG GCGGTACGCA TGCgGcgTTC
451 CACgtccAgc tGCACAACGG CAACCTAGGC GGACTCTTGA GCGGATTGCA
501 GTCGGGCGCTG GTTATGGTGT CGGGCTTTAT CGGCCTGATT GGGATGAGGA
551 TTATTTCTGT TTTTACGTCC AAACGGTTGA ACGTGCCGCA GATTCCCAGT
601 CCGAAATGGG TGGCGCAGGC TTCGCTGTGG CTACCCATGC TGACCGCCAT
651 ACTGATGGCG CACGGCGTGA TGCCTTGGCT GTCGGCGGCT TTCGCGTTTG
701 CGGCGGGCGT GATTTTACC GTACAGGTGT ACCGCTGGTG GTATAAACCC
45 751 GTATTGAAAG AACCGATGCT GTGGATTCTG TTTGCCGGCT ATCTGTTTAC
801 CGGATTGGGG CTGATTGCCG TCGGCGCGTC TTATTTCAAA CCTGCCTTCC
851 TCAATCTGGG CGTACATCTG ATCGGGGTCG GCGGTATCGG CGTGCTGACT
901 TTGGGCATGA TGGCCGTAC CGCGCTCGGT CATACGGGCA ATTCGATTTA
951 TCCGCCGCC AAAGCCGTT CCGTTGCGTT TTGGCTGATG ATGGCGGCAA
50 1001 CCGCGGTCG GTATGTTGCC GTATTTTCTT CCGGCACTGC CTACACGCAC
1051 AGCATCCGCA CGTCTCGGT TTTGTTTGCA CTCGCGCTGC TGGTGTATGC
1101 GTGGAAATAC ATTCCGTGGC TGATCCGTCC GCGTTCGGAC GGCAGGCCCG
1151 GTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 714; ORF47ng-1>:

55

```

1 MKFTKHPVWA MAFRPFYSLA ALYGALSVLL WGFYGTGTHE LSGFYWHAHE
51 MIWGYAGLVV IAFLLTAVAT WTGQPPTRGGLVGLTAFWL AARIAAFIPG
101 WGAASGILG TLFFWYGAVC MALPVIRSQNRNYVAVFAI FVLGGTHAAF

```



5  
151 HVQLHNGNLG GLLSGLQSGL VMVSGFIGLI GMRIISFFTS KRLNVPQIPS  
201 PKWVAQASLW LPMLTAILMA HGVMWPWSAA FAFAGVIFT VQVYRWYKYP  
251 VLKEPMLWIL FAGYLTGLG LIAVGASYFK PAFNLGVHL IGVGGIGVLT  
301 LGMMARTALG HTGNSIYPPP KAVPVAFWLM MAATAVRMVA VFSSGTAYTH  
351 SIRTSSVLFA LALLVYAWKY IPWLIRPRSD GRPG\*

ORF47ng-1 and ORF47-1 show 97.4% identity in 384 aa overlap:

10  
15  
20  
25  
30  
35  
40  
45

orf47-1.pep	10	20	30	40	50	60
orf47ng-1	10	20	30	40	50	60
orf47-1.pep	70	80	90	100	110	120
orf47ng-1	70	80	90	100	110	120
orf47-1.pep	130	140	150	160	170	180
orf47ng-1	130	140	150	160	170	180
orf47-1.pep	190	200	210	220	230	240
orf47ng-1	190	200	210	220	230	240
orf47-1.pep	250	260	270	280	290	300
orf47ng-1	250	260	270	280	290	300
orf47-1.pep	310	320	330	340	350	360
orf47ng-1	310	320	330	340	350	360
orf47-1.pep	370	380				
orf47ng-1	370	380				

Furthermore, ORF47ng-1 shows significant homology to an ORF from *Pseudomonas stutzeri*:

50 gnl|PID|e246540 (Z73914) ORF396 protein [Pseudomonas stutzeri] Length = 396  
Score = 155 bits (389), Expect = 5e-37  
Identities = 121/391 (30%), Positives = 169/391 (42%), Gaps = 21/391 (5%)  
Query: 7 PVWAMAFRPFYSLAALYGALSVLLWGFYGTGTHELSGFY-----WHAHEMIWGYAGLV 59  
P+W +AFRPF+ +LY L++ LW +TG GF WH HEM++G+A +  
55 Sbjct: 14 PIWRLAFRPFLLAGSLYALLAIPWVAWTGLWP--GFQPTGGWLAWHRHEMLFGFAMAI 71  
Query: 60 VIAFLLTAVATWTGQPPTRGGLVGLTAFWLAARIAAFIPGWGAAASGILGTLFFWYGAV 119  
V FLLTAV TWTGQ G LVGL A WLAAR+ ++ G AA L LF  
60 Sbjct: 72 VAGFLLTAVQTWTGQTAPSGNRLVGLAAVWLAARL-GWLFGLPAAWLAPLDLLFLVALVW 130  
Query: 120 CMALPVIRSQNRNRYVAVFAIFVLGGTHAAFXXXXXXXXXXXXXXXXXXXXXMVSGFIGL 179  
MA + + +RNY V + ++ G +V+ + L  
Sbjct: 131 MMAQMLWAVRQKRNYPIVVVLSLMLGADVLIILTGLLQGNDAQRQGVLAGLWLVAALMAL 190  
65 Query: 180 IGMRIISFFTSKRLNVPQIPSP-KWVAQASLWLPMLTAILMAHGV---MPWLSAAFAFA 234  
IG R+I FFT + L P W+ A L + A+L A GV P L F A

Sbjct: 191 IGGRVIPFFTQRLGKVDVAVKPWWLDVALLVGTGVIALLLHAFGVAMRPQPLLGLLFV-A 249

Query: 235 AGVIFTVQVYRWYKPVLPKEPMLWILFAGYLFGLGLIAGVASYF-KPAFXXXXXXXXXX 293  
 GV +++ RW+ K + K +LW L L+ + + +F A

5 Sbjct: 250 IGVGHLRLMRWYDKGIWKVGLLSLHVAMLWLVAAGFLALWHFGLLAQSSPSLHALSV 309

Query: 294 XXXXXXXXXXXMMARTALGHTGNSIYPPPKAVPVAFWLXXXXXXXXXXXXFSSGTAYTHSIR 353  
 M+AR LGHTG + P + AF L F S +

10 Sbjct: 310 GSMSGLILAMIARVTLGHTGRPLQLPAGIIG-AFVL---FNLGTAARVFLSVAPVVGGLW 365

Query: 354 TSSVLFALALLVYAWKYIPWLIRPRSDGRPG 384  
 ++V + LA +Y W+Y P L+ R DG PG

Sbjct: 366 LAAVCWTLAFALYVWRYAPMLVAARVDGHPG 396

15 Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 85

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 715>:

20 1 .. ATGCCGTCTG AAGGTTTCAGA CGGCmTCGGT GyCGGGGAay CAGAAGyGGT  
 51 AGCGCATGCC CAATGAGACT TCGTGGGTTT TGAAGCGGGT GTTTTCCAAG  
 101 CGTCCCCAGT TGTGGTAACG GTATCCGGTG TCyAArGTCA GCTTGGGyGT  
 151 GATGTCGAAa CCGACACCGG CGATGACACC AAGACCyAmG CTGCTGATrC  
 201 TGTkGCTTTC GTGATAGGsA GGTTTGyTGG kmksAsyTTG TAyrATwkkG  
 251 CCTssCwsTG kAGmGCCkTk CkyTGGTkka swGrwArTAG TCGTGGTTTy  
 25 301 TkTTyyCACC GAATGAACyT GATGTTTAAC GTGTCCGTAG GCGACGCGCG  
 351 CGCCGATATA GGGTTTGAAT TTATCGTTGA GTTTGAAATC GTAAATGGCG  
 401 GACAAGCCGA GAGAAGAAAC GGCGTGGAAG CTGCCGTTTC CCTGATGTTT  
 451 TGTTTGGGTT TCTTTGTAGT TGTTGTTTAT CTCTTCAGTA ACTTTTTTAG  
 501 TAGAAGAATT ACTTCTTTTC CATTTTCTGT AACTGGCATA ATCTGCCGCT  
 30 551 ATTCTCCAGC CGCCGAAATC ..

This corresponds to the amino acid sequence <SEQ ID 716; ORF67>:

35 1 .. MPSEGS DGXG XGEXEXVAHA QXDFVGFEAG VFQASPVVVT VSGVXXQLGX  
 51 DVETDTGDDT KTXAADXVAF VIGRFXGXLL YXXAXXXXAX XWXXXSRGF  
 101 XXHRMNL MFN VSVGDARADI GFEFIVEFEI VNGGQAERN GVEAAVSLMF  
 151 CLGFFV VVY LFSNFFSRI TFFPFSVTGI ICRYSPAEEI ..

Computer analysis of this amino acid sequence gave the following results:

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF67 shows 51.8% identity over 199 aa overlap with a predicted ORF (ORF67ng) from *N.gonorrhoeae*:

40 orf67.pep MPSEGS DGXGXGEXEXVAHAQXDFVGFEAG 30  
 orf67ng TNFEIAVLSGMTVRVIFYCARPAPVNGRLKMPSEGS DGIGIGESEAVAHQAQGFVGFEAG 146  
 90 100 110 120 130 140

45 orf67.pep VFQASPVVVTVSGVXXQLGXDVETDTGDDTKTXAADXVAFVIGRFXGXLLYXXAXXXXAX 90  
 orf67ng VFQASPVVAVAGVQGQAGRDVYAHARHRAEAQAAA VAFVIGVFLRMSVRINRNCVSI 206

50 orf67.pep XWXXXSRGFXXHRMNL MFN VSVGDARADIGFEFIVEFEIVNGGQAERNRGVEAAVSLMF 150  
 orf67ng TRVGGKSTCYFFSRIDA VSDVSVGDARTDIGFEFVFEIVNGGQAERNRGVECAVFLMF 266

```

orf67.pep      CLGFFV-----VVYLFNFFSRRITFF-PFSVTGIICRYSPAAEI      190
               | | |           :: | : | : | : | | | | | : | | | :
orf67ng        RLLVFYVKLVAAKSFIILSFQLFYVHGIFIVVFPVPTGIIRGDAPAAEVVADRHHPGVDGM 326

```

The ORF67ng nucleotide sequence <SEQ ID 717> is predicted to encode a protein comprising

5 amino acid sequence <SEQ ID 718>:

```

      1 MPSETVGSIV NVGVDES VGF SPPFPSIQHF YRFHRIHRIR LFRPPGPMQL
     51 NRHSHGSGNL GRGVWATVLS DKFPCGQVRI PACAGMTNFE IAVLSGMTVR
    101 VFYCARPAV NGGRLKMPSE GSDGIGIGES EAVAHAQRGF VGFEAGVFQA
    151 SPVVAVAGV QGQAGR DVYA HARHRAEAQA AAVAFLIGV FLRMSVRINR
    201 NCCVSITRVG GKSTCYFFSR IDAVSDVSVG DARTDIGFEF VVEFEIVNGG
    251 QAERRNGVEC AVFLMFRLLV FYVKLVAAKS FIILSFQLFY VHGIFIVVFP
    301 PVTGIIRGDA PAAEVVADRH PGVDGMRTDV SEIIAYRAYF VFAWSGWFR I
    351 IVGNAFGGVG *

```

Based on the presence of a several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 86

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 719>

```

      1 ATGTTTGCTT TTTTAGAAGC CTTTTTTGTC GAATACGGTT ATGCGGCTGT
     51 TTTTTTTGTA TTGGTCATCT GCGGTTTCGG CGTGCCGATT CCCGAGGATT
    101 TGACCTTGGT AACAGGCGGC GTGATTTTCGG GTATGGGTTA TACCAATCCG
    151 CATATTATGT TTGCAGTCGG TATGCTCGGC GTATTGGTCG GGGACGGCAT
    201 CATGTTCGCC GCCGGACGAA TTTGGGGGCA GARArTCCTA rGGTTCArAC
    251 CTATTGCGsG CATCATGACG CCGrAACGTT ATGAGCAGGT TCAGGAAAAA
    301 TTCGACAAAT ACGGTAAC TGCTTATTT GTCGCCCGTT TCCTGCCCGG
    351 TTTGAGAACG GCGGTATTG TTACAGCCGG TATCAGCCGC AAGGTTTCAT
    401 ACTTGCGTTT TATCATTATG GATGGACTGG CCGCA...

```

This corresponds to the amino acid sequence <SEQ ID 720; ORF78>:

```

      1 MFAFLEAFFV EYGYAAVFFV LVICGFGVPI PEDLTLVTGG VISGMGYTNP
     51 HIMFAVGMLG VLVGDGIMFA AGRIWGXKL XFXPIAXIMT PKRYEQVQEK
    101 FDKYGNWVLF VARFLPGLRT AVFVTAGISR KVSYLRFIIM DGLAA...

```

Further work revealed the complete nucleotide sequence <SEQ ID 721>:

```

      1 ATGTTTGCTT TTTTAGAAGC CTTTTTTGTC GAATACGGTT ATGCGGCTGT
     51 TTTTTTTGTA TTGGTCATCT GCGGTTTCGG CGTGCCGATT CCCGAGGATT
    101 TGACCTTGGT AACAGGCGGC GTGATTTTCGG GTATGGGTTA TACCAATCCG
    151 CATATTATGT TTGCAGTCGG TATGCTCGGC GTATTGGTCG GGGACGGCAT
    201 CATGTTCGCC GCCGGACGAA TTTGGGGGCA GAAAATCCTA AGGTTCAAAC
    251 CTATTGCGCG CATCATGACG CCGAAACGTT ATGAGCAGGT TCAGGAAAAA
    301 TTCGACAAAT ACGGTAAC TGCTTATTT GTCGCCCGTT TCCTGCCCGG
    351 TTTGAGAACG GCGGTATTG TTACAGCCGG TATCAGCCGC AAGGTTTCAT
    401 ACTTGCGTTT TATCATTATG GATGGACTGG CCGCACTGAT TTCCGTCCCT
    451 ATTTGGATTT ATCTGGGCGA ATACGGTGCG CACAACATCG ATTGGCTGAT
    501 GGCGAAAATG CACAGCCTGC AATCGGGTAT TTTTGTTATC TTGGGTATAG
    551 TGTGCGACCGT TGTCGCTTGG ATTTGGTGGA AAAAACGCCA ACGTATCCAG
    601 TTTTACCGCA GCAAATTGAA AGAAAAGCGG GCGCAACGCA AAGCCGCCAA
    651 GGCAGCCAAA AAAGCCGCGC AAAGCAAACA ATAA

```

This corresponds to the amino acid sequence <SEQ ID 722; ORF78-1>:

```

      1 MFAFLEAFFV EYGYAAVFFV LVICGFGVPI PEDLTLVTGG VISGMGYTNP
     51 HIMFAVGMLG VLVGDGIMFA AGRIWGXKL RFKPIARIMT PKRYEQVQEK
    101 FDKYGNWVLF VARFLPGLRT AVFVTAGISR KVSYLRFIIM DGLAALISVP
    151 IWIYLGEYGA HNIDWLMAKM HSLQSGIFVI LGIGATVVAW IWWKKRQRIQ

```

201 FYRSLKEKR AQRKAATAAK KAAQSKQ\*

Computer analysis of this amino acid sequence predicts several transmembrane domains, and also gave the following results:

Homology with the dedA homologue of *H.influenzae* (accession number P45280)

5 ORF78 and the dedA homologue show 58% aa identity in 144aa overlap:

```

Orf78: 4   FLEAFVEYGYAAVFFVLVICGFGVPIPEDLTLTVGGVISGM--GYTNPHIMFAVGMLGV 61
          FL  FF EYGY AV FVL+ICGFGVPIPED+TLV+GGVI+G+      N H+M  V M+GV
DedA:  20  FLIGFFTEYGYWAVLFVLIICGFGVPIPEDITLVSGGVIAGLYPENVNSHLMLLVSMIGV 79

Orf78: 62  LVGDGIMFAAGRIWGQXXLXFXPIAXIMTPXRYEQVQEKFDKYGNWVLFVARFLPGLRTA 121
          L GD  M+  GRI+G  L F PI  I+T  R   V+EKF +YGN VLFVARFLPGLR
DedA:  80  LAGDSCMYWLGRIYGTKILRFRIIRIVTLQRLRMVREKFSQYGNRVLFVARFLPGLRAP 139

Orf78: 122 VFVTAGISRKVSYLRFIIMDGLAA 145
          +++ +GI+R+VSY+RF+++D  AA
DedA:  140 IYMVSGITRRVSYRVFLIDFCAA 163

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF78 shows 93.8% identity over a 145aa overlap with an ORF (ORF78a) from strain A of *N*.

20 *meningitidis:*

	10	20	30	40	50	60
orf78.pep	MFAFLEAFFVEYGYAAVFVFLVICGFGVP	IPEDLTLVTGGVISGMGYTNPHIMFAVGMLG				
	:					
orf78a	MFALLEAFFVEYGYAAVFVFLVICGFGVP	IPEDLTLVTGGVISGMGYTNPHIMFAVGMLG				
	10	20	30	40	50	60
	70	80	90	100	110	120
orf78.pep	VLVGDGIMFAAGRIWGQXXLXFXPIAXIMTPXR	EYEQVQEKFDPKYGNWVLFEVARFLPGLRT				
orf78a	VLVDGDMFAAGRIWGQKILFKPIARIMTPKR	YAQVQEKFDPKYGNWVLFEVARFLPGLRT				
	70	80	90	100	110	120
	130	140				
orf78.pep	AVFVTAGISRKVSYLRFIIMDGLAA					
orf78a	AVFVTAGISRKVSYLRFLIMDGLAALISVPVWI	YLGEYGAHNIDWLMAKMHSLSQGIFIA				
	130	140	150	160	170	180

The complete length ORF78a nucleotide sequence <SEQ ID 723> is:

40	1	ATGTTTGCCC	TTTTGGAAGC	CTTTTTTGTC	GAATACGGCT	ATGCGGCCGT
	51	GTTTTTCGTT	TGGTCATCT	GCGGTTTCGG	CGTGCCGATT	CCCCGAGGATT
	101	TGACCTTGGT	AACAGCGCGC	GTGATTTTCGG	GTATGGGTTA	TACCAATCCG
	151	CATATTATGT	TTGCAGTCGG	TATGCTCGGC	GTATTGGTCG	GGGACGGCAT
	201	CATGTTTCGCC	GCCGGACGCA	TCGGGGGCA	GAAATCCTC	AAGTTCAAAC
45	251	CGATTGCGCG	CATCATGAC	CCGAAACGTT	ACGCACAGGT	TCAGGAAAAA
	301	TTCGACAAAT	ACGCCAACTG	GGTGTTATT	TGCGCTCGTT	TCCTGCCCGG
	351	TTTGCGGACT	GCCGTTTTTCG	TTACCGCCGG	CATCAGCCGC	AAAGTATCGT
	401	ATCTGCGCTT	TCTGATTATG	GACGGGCTTG	CCGCGCTGAT	TTCCGTGCC
	451	GTTTGGAATT	ACTTGGGCGA	GACGGCGCG	CACAACATCG	ATTGGCTGAT
50	501	GGCGAAAATG	CACAGCTGCG	AATCCGGCAT	CTTCATCGCA	TTGGGCGTGC
	551	TGGCGCGCGC	GCTGGCGTGG	TTCTGGTGGC	GCAAAACGCC	ACATTATCAG
	601	CTTTACCGCG	CACAATTGAG	CGAAAAACGC	GCCAAACGCA	AGGCGGAAAA
	651	GGCAGCGAAA	AAAGCGGCAC	AGAAGCAGCA	GTA	

This encodes a protein having amino acid sequence <SEQ ID 724>:

55

1	MFALLEAFFV	EYGYAAVFFV	LVICGFGVPI	PEDLTLTGTG	VISGMGYTNP
51	HIMFAVGMGL	VLVGDGIMFA	AGRIWGQKIL	KFKPIARIMT	PKRYAQVQEK
101	FDKYGNNWLF	VARFLPLGRT	AVFVTAGISR	KVSYLRFLIM	DGLAALISVP

151 VWIYLGEYGA HNIDWLMAMK HSLSQSGIFIA LGVLAAALAW FWWRKRRHYQ  
 201 LYRAQLSEKR AKRKAKEAAK KAAQKQQ\*

ORF78a and ORF78-1 show 89.0% identity in 227 aa overlap:

		10	20	30	40	50	60
5	orf78a.pep	MFALLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNPHIMFAVGMLG					
	orf78-1	MFAFLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNPHIMFAVGMLG					
		10	20	30	40	50	60
10	orf78a.pep	VLVGDGIMFAAGRIWGQKILKFKPIARIMTPKRYAQVQEKFDKYGNWVLFVARFLPGLRT					
	orf78-1	VLVGDGIMFAAGRIWGQKILKFKPIARIMTPKRYEQVQEKFDKYGNWVLFVARFLPGLRT					
		70	80	90	100	110	120
15	orf78a.pep	AVFVTAGISRKVSYLRLIMDGLAALISVPVWIYLGEYGAHNIDWLMAMKHSLSQSGIFIA					
	orf78-1	AVFVTAGISRKVSYLRFIIMDGLAALISVPIWIYLGEYGAHNIDWLMAMKHSLSQSGIFVI					
		130	140	150	160	170	180
20	orf78a.pep	LGVLAAALAWFWWRKRRHYQLYRAQLSEKRAKRAEKAAKKAAQKQQX					
	orf78-1	LGIGATVVAWIWKKRQRIQFYRSKLKEKRAQRKAAKAAKAAQSKQX					
		190	200	210	220		
25	orf78a.pep	LGVLAAALAWFWWRKRRHYQLYRAQLSEKRAKRAEKAAKKAAQKQQX					
	orf78-1	LGIGATVVAWIWKKRQRIQFYRSKLKEKRAQRKAAKAAKAAQSKQX					
		190	200	210	220		

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF78 shows 97.4% identity over 38 aa overlap with a predicted ORF (ORF78ng) from *N. gonorrhoeae*:

30	orf78.pep	XXLFXPIAXIMTPXRYEQVQEKFDKYGNWVLFVARFLPGLRTAVFVTAGISRKVSYLRF	137
	orf78ng	YPVLFVARFLPGLRTAVFVTAGISRKVSYLRF	32
35	orf78.pep	IIMDGLAA	145
	orf78ng	LIMDGLAALISVPVWIYLGEYGAHNIDWLMAMKHSLSQSGIFIALGVLAALAWFWWRKRR	92

The ORF78ng nucleotide sequence <SEQ ID 725> is predicted to encode a protein comprising amino acid sequence <SEQ ID 726>:

40	1	..YPVLFVARFL PGLRTAVFVT AGISRKVSYL RFLIMDGLAA LISVPVWIYL
	51	GEYGAHNIDW LMAKMHSLSQ GIFIALGVLA AALAWFWWRK RRHYQLYRAQ
	101	LSEKRAKRKA EKAAKKAAQK QQ*

Further work revealed the complete gonococcal nucleotide sequence <SEQ ID 727>:

45	1	atgtttgccc tttTggaagc CTTTTTGTG	GAAatcgCt atgcGGCCGT
	51	GTTTTTCGTT TTGGTCATCT GCGGTTTCGG	CGTGCCGATT CCCGAAGATT
	101	TGACCTTGGT AACGGCGCGC GTGATTCGG	GTATGGGTTA TACCAATCCG
	151	CATATTATGT TTGCGGTCGG TATGCTCGGC	GTGTTGGCGG GCGACGGCGT
	201	GATGTTTGCC GCCGACGCA TCTGGGGGCA	GAAAATCCTC AAGTTCAAAC
	251	CGATTGCGCG CATCATGACG CCGAAACGTT	ACGCGCAGGT TCAGGAAAAA
50	301	TTCGACAAAT ACGGCAACTG GGTCTGTGT	GTCGCCCGTT TCCTGCCGGG
	351	TTTGCGGACT GCCGTTTTTCG TTACCGCCGG	CATCAGCCGC AAAGTATCGT
	401	ATCTGCGCTT TCTGATTATG GACGGGCTGG	CCGCGCTGAT TTCCGTGCC
	451	GTTTGGATTT ACTTGGGCGA GTACGGCGCG	CACAACATCG ATTGGCTGAT
	501	GGCGAAAATG CACAGCCTGC AATCGGGCAT	CTTCATCGCA TTGGGCGTGC
55	551	TGGCGGCGGC GCTGGCGTGG TTCTGGTGGC	GCAAACGCCG ACATTATCAG
	601	CTTTACCGCG CACAATTGAG CGAAAAACGC	GCCAAACGCA AGGCGGAAAA
	651	GGCAGCGAAA AAAGCGGCAC AGAAGCAGCA	GTAA

This corresponds to the amino acid sequence <SEQ ID 728; ORF78ng-1>:

5

1	MFALLEAFFV	EYGYAAVFFV	LVICFGVPI	PEDLT <del>LV</del> TGG	VISGMGYTNP
51	HIMFAVGLG	<u>VL</u> AGDGV <del>M</del> FA	AGRIWQKIL	KFKPIARIMT	PKRYAQVQEK
101	FDKYGNWVLF	VARELPGLRT	AVFVTAGISR	KVSYLRFLIM	DGLAALISVP
151	WVIYLGEYA	HNIDWLMAKM	<u>HSLQSGIFIA</u>	LGVLAAALAW	FWWRKRHRHY
201	LYRAOLSEKR	AKRKA <del>E</del> AKA	KAAQKQ*		

ORF78ng-1 and ORF78-1 show 88.1% identity in 227 aa overlap:

		10	20	30	40	50	60
10	orf78-1.pep	MFAFLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNPHIMFAVGMLG					
	orf78ng-1	MFALLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNPHIMFAVGMLG					
		10	20	30	40	50	60
15	orf78-1.pep	VLVGDGIMFAAGRIWGQKILRFKPIARIMTPKRYEQVQEKFDKYGNWVLFVARFLPGLRT					
	orf78ng-1	VLAGDGMFAAGRIWGQKILKFKPIARIMTPKRYAQVQEKFDKYGNWVLFVARFLPGLRT					
		70	80	90	100	110	120
20	orf78-1.pep	AVFVTAGISRKVSYLRFIIMDGLAALISVPIWIYLGEYGAHNIDWLMAKMHSLSQSGIFVI					
	orf78ng-1	AVFVTAGISRKVSYLRFIIMDGLAALISVPVWIYLGEYGAHNIDWLMAKMHSLSQSGIFIA					
		130	140	150	160	170	180
25	orf78-1.pep	LGIGATVVAWIWKKRQRIQFYRSKLKEKRAQRKAAKAAKAAQSKQX					
	orf78ng-1	LGVLAAALAWFWWRKRHRHYQLYRAQLSEKRAKRKAekaakkaAQKQXX					
		190	200	210	220		
30	orf78-1.pep	LGIGATVVAWIWKKRQRIQFYRSKLKEKRAQRKAAKAAKAAQSKQX					
	orf78ng-1	LGVLAAALAWFWWRKRHRHYQLYRAQLSEKRAKRKAekaakkaAQKQXX					
		190	200	210	220		

Furthermore, orf78ng-1 shows homology to the dedA protein from *H.influenzae*:

sp|P45280|YG29\_HAEIN HYPOTHETICAL PROTEIN HI1629 >gi|1073983|pir||D64133 dedA  
protein (dedA)\_homolog - Haemophilus influenzae (strain Rd KW20)  
>gi|1574476 (U32836) dedA protein (dedA) [Haemophilus influenzae] Length = 212  
Score = 223 bits (563), Expect = 7e-58  
Identities = 108/182 (59%), Positives = 140/182 (76%), Gaps = 2/182 (1%)

Query: 5 LEAFFVEYGYAAVFFVLICGFGVPIPEDLTLVTGGVISGM--GYTNPHIMFAVGMGLVL 62  
L FF EYGY AV FVL+ICGFGVPIPED+TLV+GGVI+G+ N H+M V M+GVL  
Spict: 21 LIGFFTEYGYWAVLFVLIICGFGVPIPEDITLVSGGVIAGLYPENVNSHLMLLVSMIGVL 80

Query: 63 AGDGVMAAGRIWGQKILFKPIARIMTPKRYAQVQEKFDKYGNWVLFVARFLPGLRTAV 122  
AGD M+ GRI+G KIL+F+PI RI+T +R V+EKF +YGN VLFVARFLPGLR +  
Sbjct: 81 AGDSCMYWLGRIYGTKILRFRPIRRIVTLQRLRMVREKFSOYGNRVLFVARFLPGLRAPI 140

Query: 123 FVTAGISRKVSYRLFLIMDGLAALISVPVWIYLGEGAHNIDWLMAKMHSLSQSGIFIALG 182  
 ++ +GI+R+VSY+RF+++D AA+ISVP+WIIYLG E GA N+DWL ++ Q I+I +G  
 Sbjct: 141 YMVSGITRRVSYRVFLIDFCAAIIISVPIWIYLGELGAKNLDWLHTQIQKGQIVIIYIFIG 200

```
Query: 183 VL 184
        L
Sbjct: 201 YL 202
```

Based on this analysis, including the presence of putative transmembrane domains, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

**Example 87**

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 729>:

```

1  ATGAAAAAAT TATTGGCGGC CGTGATGATG GCAGGTTTGG CAGGCGCGGT
51  TTCCGCCGCC GGAGTCCACG TTGAGGACGG CTGGGCGCGC ACCACCGTCG
101 AAGGTATGAA AATAGGCGGC GCGTTCATGA AAATCCACAA CGACGAAGCC
151 AAACAAGACT TTTTGCTCGG CGGAAGCAGC CCCGTTGCCG ACCGCGTCGA
201 AGTGCATACC CACATCAACG ACAACGGCGT GATGCGGATG CGCGAAGTCG
251 AAGGCGGCGT GCCTTTGGAA GCGAAATCCG TTACCGAACT CAAACCCGGC
301 AGCTATCATG TGATGTTTAT GGGTTTGAAA AAACAATTAA AAGAGGGCGA
10 351 TAAAATTCCT GTTACCCTGA AATTTAAAAA CGCCAAAGCG CAAACCGTCC
401 AACTGGAAGT CAAATCGCG CCGATGCCGG CAATGAACCA C...

```

This corresponds to the amino acid sequence <SEQ ID 730; ORF79>:

```

1  MKKLLAAVMM AGLAGAVSAA GVHVEDGWAR TTVEGMKIGG AFMKIHNDDEA
51  KQDFLLGGSS PVADRVEVHT HINDNGVMRM REVEGGVPLE AKSVTELKPG
15 101 SYHVMFMGLK KQLKEGDKIP VTLKFKNAKA QTVQLEV KIA PMPAMNH..

```

Further work revealed the complete nucleotide sequence <SEQ ID 731>:

```

1  ATGAAAAAAT TATTGGCGGC CGTGATGATG GCAGGTTTGG CAGGCGCGGT
51  TTCCGCCGCC GGAGTCCACG TTGAGGACGG CTGGGCGCGC ACCACCGTCG
20 101 AAGGTATGAA AATAGGCGGC GCGTTCATGA AAATCCACAA CGACGAAGCC
151 AAACAAGACT TTTTGCTCGG CGGAAGCAGC CCCGTTGCCG ACCGCGTCGA
201 AGTGCATACC CACATCAACG ACAACGGCGT GATGCGGATG CGCGAAGTCG
251 AAGGCGGCGT GCCTTTGGAA GCGAAATCCG TTACCGAACT CAAACCCGGC
301 AGCTATCATG TGATGTTTAT GGGTTTGAAA AAACAATTAA AAGAGGGCGA
25 351 TAAAATTCCT GTTACCCTGA AATTTAAAAA CGCCAAAGCG CAAACCGTCC
401 AACTGGAAGT CAAATCGCG CCGATGCCGG CAATGAACCA CGGTATCATC
451 CACGCGGAAG CGCATCAGCA CTAA

```

This corresponds to the amino acid sequence <SEQ ID 732; ORF79-1>:

```

1  MKKLLAAVMM AGLAGAVSAA GVHVEDGWAR TTVEGMKIGG AFMKIHNDDEA
51  KQDFLLGGSS PVADRVEVHT HINDNGVMRM REVEGGVPLE AKSVTELKPG
30 101 SYHVMFMGLK KQLKEGDKIP VTLKFKNAKA QTVQLEV KIA PMPAMNHGHH
151 HGEAHQH*

```

Computer analysis of this amino acid sequence revealed a putative leader peptide and also gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

35 ORF79 shows 94.6% identity over a 147aa overlap with an ORF (ORF79a) from strain A of *N.meningitidis*:

```

              10      20      30      40      50      60
orf79.pep    MKKLLAAVMMAGLAGAVSAAGVHVEDGWARTTVEGMKIGGAFMKIHNDDEAKQDFLLGGSS
40 orf79a     MKKLLAAVMMAGLAGAVSAAGIHVEDGWARTTVEGMKMGGAFMKIHNDDEAKQDFLLGGSS
              10      20      30      40      50      60
              70      80      90     100     110     120
orf79.pep    PVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIP
45 orf79a     PVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGXKKQLKXGDKIP
              70      80      90     100     110     120
              130     140
orf79.pep    VTLKFKNAKAQTVQLEV KIA PMPAMNH
50 orf79a     VTLKFKNAKAQTVQLEVKTAPMSAMDHGHHHGEAHQH
              130     140     150

```

The complete length ORF79a nucleotide sequence <SEQ ID 733> is:

```

      1  ATGAAANAAC  TATTGGCAGC  CGTGATGATG  GCAGGTTTGG  CAGGCGCGGT
     51  TTCCGCCGCC  GGAATCCACG  TTGAGGACGG  CTGGGCGCGC  ACCACCGTCG
    101  AAGGTATGAA  AATGGGCGGC  GCGTTCATGA  AAATCCACAA  CGACGAAGCC
    151  AAACAAGACT  TTTTGCTCGG  CGGAAGCAGC  CCTGTTGCCG  ACCGCGTCGA
    201  AGTGCATACC  CATATCAATG  ATAACGGTGT  GATGCGGATG  CGCGAAGTCG
    251  AAGGCGGCGT  GCCTTTGGAG  GCGAAATCCG  TTACCGAACT  CAAACCCGGC
    301  AGCTATCATG  TCATGTTTAT  GGGTNTGAAA  AAACAATTAA  AAGANGGCCA
    351  CAAGATTCCC  GTTACCCTGA  AATTAAAAAA  CGCCAAAGCA  CAAACCGTCC
    401  AACTGGAAGT  CAAAACCGCG  CCGATGTCGG  CAATGGACCA  CGGTCATCAC
    451  CACGGCGAAG  CGCATCAGCA  CTAA
  
```

This encodes a protein having amino acid sequence <SEQ ID 734>:

```

      1  MKXLLAAVMM  AGLAGAVSAA  GIHVEDGWAR  TVEGMKMG  AFMKIHND  EA
     51  KQDFLLGGSS  PVADRVVHT  HINDNGVMRM  REVEGGVPLE  AKSVTELKPG
    101  SYHVMFMGXK  KQLKXGDKIP  VTLKFKNAKA  QTVQLEVKTA  PMSAMDHGHH
    151  HGEAHQH*
  
```

ORF79a and ORF79-1 show 94.9% identity in 157 aa overlap:

```

      10      20      30      40      50      60
    orf79a.pep  MKXLLAAVMMAGLAGAVSAAGIHVEDGWARTTVEGMKMGGA FMKIHND EAKQDFLLGGSS
    orf79-1     MKKLLAAVMMAGLAGAVSAAGVHVEDGWARTTVEGMKIGGA FMKIHND EAKQDFLLGGSS
      10      20      30      40      50      60
    orf79a.pep  PVADRVVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGXKKQLKXGDKIP
    orf79-1     PVADRVVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIP
      70      80      90     100     110     120
    orf79a.pep  VTLKFKNAKAQTVQLEVKTAPMSAMDHGHHHGEAHQH
    orf79-1     VTLKFKNAKAQTVQLEVKTAPMPAMNHGHHHGEAHQH
      70      80      90     100     110     120
    orf79a.pep  VTLKFKNAKAQTVQLEVKTAPMSAMDHGHHHGEAHQH
    orf79-1     VTLKFKNAKAQTVQLEVKTAPMPAMNHGHHHGEAHQH
      130     140     150
    orf79a.pep  VTLKFKNAKAQTVQLEVKTAPMSAMDHGHHHGEAHQH
    orf79-1     VTLKFKNAKAQTVQLEVKTAPMPAMNHGHHHGEAHQH
      130     140     150
  
```

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF79 shows 96.1% identity over 76 aa overlap with a predicted ORF (ORF79ng) from *N.gonorrhoeae*:

```

    orf79.pep  FMKIHND EAKQDFLLGGSSPVADRVVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGS  101
    orf79ng    INDNGVMRMREVKGVPLEAKSVTELKPGS  30
    orf79.pep  YHVMFMGLKKQLKEGDKIPVTLKFKNAKAQTVQLEVKTAPMPAMNH  147
    orf79ng    YHVMFMGLKKQLKEGDKIPVTLKFKNAKAQTVQLEVKTAPMSAMNHGHHHGEAHQH  86
  
```

An ORF79ng nucleotide sequence <SEQ ID 735> was predicted to encode a protein comprising amino acid sequence <SEQ ID 736>:

```

      1  ..INDNGVMRM  EVKGGVPLEA  KSVTELKPGS  YHVMFMGLKK  QLKEGDKIPV
     51  TLKFKNAKAQ  TVQLEVKTAP  MSAMNHGHHH  GEAHQH*
    101  aaggtATgaa  aatggGCGGC  GCgttCATga  aaATCCACAA  CGACGaaGcc
    151  atacaaGACt  ttgtgcTCgg  CGGaagcatg  cccgttgccg  accgcGTGCA
    201  AGTGCAAtaca  cacATCAACG  ACAACGGCGT  GATGCGTATG  CGCGAAGTCA
  
```

Further work revealed the complete gonococcal DNA sequence <SEQ ID 737>:



5  
251 AAGGCGGCGT GCCTTTGGAG GCGAAATCCG TTACCGAACT CAAACCCGGC  
301 AGCTATCACG TGATGTTTAT GGGTTTGAAG AAACAACCTGA AAGAGGGCGA  
351 CAAGATTCCC GTTACCCTGA AATTTAAAAA CGCCAAAGCG CAAACCGTCC  
401 AACTGGAAGT CAAAACCGCG CCGATGTCGG CAATGAACCA CGGTCATCAC  
451 CACGGCGAAG CGCATCAGCA CTAA

This corresponds to the amino acid sequence <SEQ ID 738; ORF79ng-1>:

10  
1 MKKLLAAVMM AGLAGAVSAA GVHVEDGWAR TTVEGMKMG AFMKIHND  
51 IQDFVLGGSM PVADRVEVHT HINDNGVMRM REVKGGVPLE AKSVTELKPG  
101 SYHVMFMGLK KQLKEGDKIP VTLKFKNAKA QTVQLEVKTA PMSAMNHGHH  
151 HGEAHQH\*

ORF79ng-1 and ORF79-1 show 95.5% identity in 157 aa overlap:

15  
10 20 30 40 50 60  
orf79-1.pep MKKLLAAVMMAGLAGAVSAAGVHVEDGWAR TTVEGMKMG AFMKIHND EAKQDFLLGGSS  
|||||  
orf79ng-1 MKKLLAAVMMAGLAGAVSAAGVHVEDGWAR TTVEGMKMG AFMKIHND EAKQDFVLGGSM  
10 20 30 40 50 60  
20  
70 80 90 100 110 120  
orf79-1.pep PVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIP  
|||||  
orf79ng-1 PVADRVEVHTHINDNGVMRMREV KGGVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIP  
70 80 90 100 110 120  
25  
130 140 150  
orf79-1.pep VTLKFKNAKAQTVQLEV K I A P M P M N H G H H G E A H Q H X  
|||||  
orf79ng-1 VTLKFKNAKAQTVQLEV K T A P M S A M N H G H H G E A H Q H X  
130 140 150

Furthermore, ORF79ng-1 shows significant homology to a protein from *Aquifex aeolicus*:

30  
gi|2983695 (AE000731) putative protein [Aquifex aeolicus] Length = 151  
Score = 63.6 bits (152), Expect = 6e-10  
Identities = 38/114 (33%), Positives = 58/114 (50%), Gaps = 1/114 (0%)  
35  
Query: 24 VEDGWAR TTVEGMKMG AFMKIHND EAKQDFVLGGSM PVADRVEVHTHINDNGVMRMREV 83  
V+ W G M I N+ D+++G +A RVE+H + +N V +M  
Sbjct: 27 VKHPWVMEPPPGFNTTMMGMII VNEGDEPDYLGAKTDIAQRVELHKTVIENDVAKMVPQ 86  
40  
Query: 84 KGGVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIPVTLKFKNAKAQTVQLEV 137  
+ + + K E K YHVM +GLKK++KEGDK+ V L F+ + TV+ V  
Sbjct: 87 ER-IEIPPGKVFEFKHHGYHVMII GLKKRIKEGDKVKVELIFEKSGKITVEAPV 139

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

45 ORF79-1 (15.6kDa) was cloned in the pET vector and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 18A shows the results of affinity purification of the His-fusion protein. Purified His-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result) and FACS analysis (Figure 18B) These experiments confirm that ORF79-1 is a surface-exposed protein, and that it is a useful immunogen.

5	1	ATGACGGTAA	CTGCGGCCGA	AGGCGGCAAA	GCTGCCAAGG	CGTAAAAAAA
	51	ATATCTGATT	ACGGGGATTT	TGGTCTGGCT	GCCGATTGCG	GTAACCGGTT
	101	GGGTGGTTTC	CTATATCGTT	TCCGCGTCCG	ATACGCTGCT	CACCTCGTTC
	151	CCGAAGCAAT	GGCGGCCGCA	ATATGTTTTG	GGGTTTAATA	TCCCGGGGCT
	201	GGGCGTTATC	GTTGCCATTG	CCGTATTGTT	TGTAAACCGA	TGTGTTGCCG
10	251	CCAACGTATT	GGGTGCGCAG	ATCCTCGCCG	CGTGGGACAG	CGTGTGGGG
	301	CGGATTCCGG	TTGTGAAATc	CATCTATTCT	AGTGTGAAAA	AAGTATCCGA
	351	ATaagTGCTG	TCCGACAGCA	GCCGTTCTGTT	TAAACACGCC	GTACTCGTGC
	401	CGTTTCCCCA	GCCCGGTATT	TGGACGATyG	CTTTCTGTGTG	AGGGCAGTGC
	451	TCGAATGCGG	TTAAGGCCCG	ATTGCCGAAs	GACGGCGATT	ATCTTTCCGT
15	501	GTAATGTCGG	ACCACGCCGA	ATCCGACCCG	CGGTTACTAT	ATTATGGTAA
	551	AGAAAGCGA	TGTGCGCGAA	CTCGATATGA	CCGTGGACGA	AsCATTGAAA
	601	TATGTGATTT	CGCTGGGTAT	GGTCATCCCT	GACGACCTGC	CCGTCAAAAC
	651	ATTGGCAsGA	CCTATGCCGT	CTGAAAGGCG	GGATTTTGCC	GAACAACAAT
	701	AA				

20

1	MTVTAAEGGK	AAKALKKKYLI	TGILVWLPIA	VTVWVVSIV	SASDQLVNL
51	PKQWRPQYVL	GFNIPGLGVI	VAIAVLFTGT	LFAANVLGRQ	ILAAWDSILG
101	RIPVVKSIYS	SVKKVSEYVL	SDSSRSFKTP	VLVPPFPQGI	WTAFAVSGQV
151	SNAVKAALPX	DGDYLSVYVP	TFPNFTGGYI	IMVKKSDVRE	LDMSVDEXLK
201	YVTSLGMVIP	DDLVPKTLAX	PMPSEKADLP	EQQ*	

	1	ATGACGGAAC	nTGCGGCCGA	AGGCGGCAAA	GCTGCCAArG	CGTAAAAAAA
	51	ATATCTGATT	ACGGGCATTT	TGGTCTGGCT	GCCGATTGCG	GTACCGGTTT
	101	GGGTGGTTTC	CTATATCGTT	TCCGCGTCCG	ATCAGCTCGT	CACCTGCTGT
30	151	CCGAAGCAAT	GGCGGCCGCA	ATATGTTTTG	GGGTTTAAAT	TCCCGGGGCT
	201	GGGCGTTATC	GTTGCCATTG	CCGTATTGTT	TGTAACCGGA	TTGTTTGCCG
	251	CCAACGTATT	GGGTCCGGAC	ATCTCGCCG	CGTGGGACAG	CCGTGTTGGG
	301	CGGATCCGG	TGTGAAATC	CATCTATTCT	AGTGTGAAAA	AAGTATCCGA
	351	ATCGCTGCTG	TCCGACAGCA	GCCGTTCTGT	TAAACAGCCG	GTACTCGTGT
35	401	CGTTTCCCCA	GCCCGGTATT	TGGACGATTG	CTTTCTGTCT	AGGGCAGGTG
	451	TCGAATGCCG	TTAAGGCCGC	ATTGCCGAAG	GACGGCGATT	ATCTTTCCGT
	501	GTATGTTCCG	ACCACGCCGA	ATCCGACCCG	CGGTTACTAT	ATTATGGTAA
	551	AGAAATAGCA	TGTGCGCGAA	CTCGATATGA	CGGTGGACGA	AGCATGAAA
	601	TATGTGATTT	CGCTGGGTAT	GGTCAATCCCT	GACGACCTGC	CCGTCAAAC
40	651	ATTGGCAGGA	CCTATGCCGT	CTGAAAGGCG	GGATTTGCC	GAACAACAAT
	701	AA				

45

1	MTEXAAEGGK	AAKALKKKYLI	TGILVWLPIA	VTVWVVSIV	SASDQLVNL
51	PKQWRPQYVL	GFNIPGLGVI	VAIAVLFTGT	LFAANVLGRQ	ILAAWDSLIG
101	RIPVVKSIYS	SVKKVSESL	SDSSRSFKTP	VLVFPFPQGI	WTIAFVSGQV
151	SNAVKAALPK	DGDYLSVYVP	TTPNPTGGY	IMVKKSDVRE	LDMSVDEALK
201	YVISLGMVIP	DDLVPKTLAG	PMPSEKADLP	EQQ*	

Homology with a predicted ORF from *N.meningitidis* (strain A)

50    *meningitidis*:

                          10                  20                  30                  40                  50                  60

orf98.pep    MTVTAAEGGKAAKALKKKYLITGILVWLPIAVTVWVVSYSIVSASDQLVNLLPKQWRPQYVL

5	orf98a	MTEPAAEGGKAALKKKYLITGILVWLPIAVTVWVVSIVSASDQLVNLLPKQWRPQYVL	10	20	30	40	50	60
			70	80	90	100	110	120
10	orf98.pep	GFNIPGLGVIVAIIVLFVTGLFAANVLGRQILAAWDSLGRIPVVKSIYSSSVKKVSEYVL						
	orf98a	GFNIPGLGVIVAIIVLFVTGLFAANVLGRQILAAWDSLGRIPVVKSIYSSSVKKVXSLL	70	80	90	100	110	120
15	orf98.pep	SDSSRSFKTPVLVFPFQPGIWTIAFVSGQVSNVKAALPKDGDYLSVYVPTTNPNTGGYY						
	orf98a	SDSSRSFKTPVLVFPFQSGIWTIAFVSGQVSNVKAALPKDGDYLSVYVPTTNPNTGGYY	130	140	150	160	170	180
20	orf98.pep	IMVKKSDVRELDMSVDEALKYVISLGMVIPDDLVPVKTLAXPMPSEKADLPEQQX						
	orf98a	IMVKKSDVRELDMSVDEALKYVISLGMVIPDDLVPVKTLAGPMPSEKADLPEQQX	190	200	210	220	230	

The complete length ORF98a nucleotide sequence <SEQ ID 743> is:

25	1	ATGACGGAAC	CTGCGGCCGA	AGGCGGCAAA	GCTGCCAAGG	CGTTAAAAAA
	51	ATATCTGATT	ACGGGCATTT	TGGTCTGGCT	GCCGATTGCG	GTAACGGTTT
30	101	GGGTGGTTTC	CTATATCGTT	TCCGCGTCCG	ATCAGCTCGT	CACCTGCTG
	151	CCGAAGCAAT	GGCGGCCGCA	ATATGTTTGT	GGGTTTAATA	TCCCGGGGCT
35	201	GGCGTTATC	GTGCCCATTG	CCGTATTGTT	TGTAACCGGA	TTATTTGCCG
	251	CAAACGTATT	GGGCCGGCAG	ATTCTTGCCG	CGTGGGACAG	CTTGTGGGG
40	301	CGGATTCCGG	TGTGAAGTC	CATCTATTCG	AGTGTGAAA	AAGTATCCGA
	351	NTCGTTGCTG	TCCGACAGCA	GCCGTTCTGT	TAAAACACCA	GTACTCGTGC
45	401	CGTTTCCCCA	ATCGGGTATT	TGGACAATCG	CATTCGTGTC	CGGTCAGGTG
	451	TCGAATGCGG	TTAAGGCCGC	ATTGCCGAAG	GACGGCGATT	ATCTTCCGT
50	501	GTATGTTCCG	ACCACGCCGA	ATCCGACCGG	CGGTTACTAT	ATTATGGTAA
	551	AGAAAAGCGA	TGTGCGCGAA	CTCGATATGA	GCGTGGACGA	AGCGTTGAAA
55	601	TATGTGATTT	CGCTGGGTAT	GGTCATCCCT	GACGACCTGC	CCGTCAAAAC
	651	ATTGGCAGGA	CCTATGCCGT	CTGAAAAGGC	GGATTTGCCC	GAACAACAAT
	701	AA				

This encodes a protein having amino acid sequence <SEQ ID 744>:

40	1	MTEPAAEGGK	AAKALKKYL	ITGILVWLPIA	VTWVVSIV	SASDQLVNLL
	51	PKQWRPQYVL	GFNIPGLGVI	VAIVLFVTG	LFAANVLGRQ	ILAAWDSLGR
45	101	RIPVVKSIYS	SVKKVXSLL	SDSSRSFKTP	VLVFPQSGI	WTIAFVSGQV
	151	SNAVKAALPK	DGDYLSVYVP	TPNPNTGGYY	IMVKKSDVRE	LDMSVDEALK
	201	YVISLGMVIP	DDLVPVKTLA	GMPSEKADLP	EQQ*	

ORF98a and ORF98-1 show 98.7% identity in 233 aa overlap:

45	orf98a.pep	MTEPAAEGGKAALKKKYLITGILVWLPIAVTVWVVSIVSASDQLVNLLPKQWRPQYVL	10	20	30	40	50	60
	orf98-1	MTEXAAEGGKAALKKKYLITGILVWLPIAVTVWVVSIVSASDQLVNLLPKQWRPQYVL	10	20	30	40	50	60
50	orf98a.pep	GFNIPGLGVIVAIIVLFVTGLFAANVLGRQILAAWDSLGRIPVVKSIYSSSVKKVXSLL						
	orf98-1	GFNIPGLGVIVAIIVLFVTGLFAANVLGRQILAAWDSLGRIPVVKSIYSSSVKKVSESLL	70	80	90	100	110	120
55	orf98a.pep	SDSSRSFKTPVLVFPFQSGIWTIAFVSGQVSNVKAALPKDGDYLSVYVPTTNPNTGGYY						
	orf98-1	SDSSRSFKTPVLVFPFQPGIWTIAFVSGQVSNVKAALPKDGDYLSVYVPTTNPNTGGYY	130	140	150	160	170	180
60	orf98a.pep	IMVKKSDVRELDMSVDEALKYVISLGMVIPDDLVPVKTLAGPMPSEKADLPEQQX						
			190	200	210	220	230	

orf98-1 IMVKKSDVRELDMSVDEALKYVISLGMVIPDDL PVKTLAGMPSEKADLPEQQX  
190 200 210 220 230

5 Homology with a predicted ORF from *N.gonorrhoeae*

ORF98 shows 95.3% identity over a 233 aa overlap with a predicted ORF (ORF98ng) from *N.gonorrhoeae*:

		10	20	30	40	50	60	
10	orf98.pep	MTVTAAEGGKAALKKKYLITGILVWLPIAVTVVVS	YIVSASDQLVNLLPKQWRPQYVL	60				
	orf98ng	MTEPAAEGGKAALKKKYLITGILVWLPIAVTVVVS	YIVSASDQLVNLLPKQWRPQYVL	60				
15	orf98.pep	GFNIPGLGVIVAI	IAVL	FVTGLFAANVLGRQILA	AWDSSLGRIPVVKSIYSSVKVSEYVL	120		
	orf98ng	GFNIPGLGVIVAI	IAVL	FVTGLFAANVLGRQILA	AWDSSLXRI	PVVKSIYSSVKVSESL	120	
20	orf98.pep	SDSSRSFKTPVLVPFPQPGIWTIAFVSGQV	SNAVKAAL	EPDGDYLSVYVPTT	PNPTGGYY	180		
	orf98ng	SDSSRSFKTPVLVPFPQSGIWTIAFVSGQV	SNAVKAAL	PQDGDYLSVYVPTT	PNPTGGYY	180		
	orf98.pep	IMVKKSDVRELDMSVDE	XLKYVISLGMVIPDDLPVK	TLAXPMPSEKADLPEQQ	233			
	orf98ng	IMVKKSDVRELDMSVDE	ALKYVISLGMVIPDDLPVK	TLAGPMPPEKAELPEQQ	233			

The complete length ORF98ng nucleotide sequence <SEQ ID 745> is predicted to encode a protein having amino acid sequence <SEQ ID 746>:

30

1	MTEPAAEGGK	AAKALKKYLI	TGILVWLPIA	VTWVVVSIV	SASDQLVNL
51	PKQWRPQYVL	GFNIPGLGVI	VAIAVLFTG	LFAANVLGRQ	ILAAWDSLLX
101	RIPVVKSIYS	SVKKVSESL	SDSSRSFKTP	VLVPFPQSGI	WTIAFVSGQV
151	SNAVKAALPQ	GDGYSVYVP	TTPNPTGGY	IMVKSDVRE	LDMSVDEALK
201	YVISLGMVIP	DDLVPKTLAG	PMPPEKAELP	EQO*	

Further work revealed the complete nucleotide sequence <SEQ ID 747>:

	1	ATGACGGAAC	CTGCGGCCGA	AGGCGGCAAA	GCTGCCAAGG	CGTTAAAAAA
	51	ATATCTGATT	ACAGGCATTT	TGGTCTGGCT	GCCGATTGCG	GTAAACGGTTT
35	101	GGGTGGTTTT	CTATATCGTT	TCCGCGTCCG	ACACAGTGTG	CAACCTGCTG
	151	CGGAAGCAAT	CGCGGCCCA	ATATGTTTTG	GGGTTTAATA	TCCCCGGGCT
	201	CGGCGTTATT	GTTGCCATTG	CCGTATTGTT	TGTAACCGGA	TTATTTGCCG
	251	CAAACGTGTT	GGGCCGGCAG	ATTCTTGCCG	CGTGGGACAG	CCTGTTgggg
	301	cggaTTCGGG	TTGTCAAATC	CATCTATTCT	AGTGTGAAAA	AAGTATCCGA
40	351	ATCGCTGCTG	TCCGACAGCA	GCCGTTCTGT	TAAACGCCCG	GTACTCGTGC
	401	CGTTTCCCCA	ATCGGGTATT	TGGACAATCG	CATTCTGTGC	CGGTCAGGTG
	451	TCGAATGCGG	TTAAGGCCGC	ATTGCCGCG	GATGGCGATT	ATCTTTCCGT
	501	GTATGTCCCG	ACCACGCCCA	ACCGACCGAG	CGGTTACTAT	ATTATGGTAA
	551	AGAAAAGCGA	TGTGCGCGAA	CTCGATATGA	GCGTGGACGA	AGCGTTGAAA
45	601	TATGTGATTT	CGCTGGGTAT	GGTCATCCCT	GACGACCTGC	CCGTCAA AAC
	651	ATTGGCAGGA	CCTATGCCGC	CTGAAAAGGC	GGAGTTGCCC	GAACAACAAT
	701	AA				

This corresponds to the amino acid sequence <SEQ ID 748; ORF98ng-1>:

50

1	MTEPAAEGGK	AAKALKKYLI	TGILVWLPIA	VTWVVSIV	SASDQLVNL
51	PKQWRPQYVL	GFNIPGLGVI	VAIAVLFTG	LFAANVLGRQ	ILAAWDSL
101	RIPVVKSIYS	SVKKVSESL	SDSSRSFKTP	VLVFPQSGI	WTIAFVSGQV
151	SNAVKAAALPQ	DDYLSVYVP	TPNPDTGGY	IMVKSDVRE	LDMSVDEALK
201	YVISLGMVIP	DGLPVKTLAG	PMPEKAELP	EQQ*	

ORF98ng-1 and ORF98-1 show 97.9% identity in 233 aa overlap:

55 orf98-1.pep MTEXAAEGGKAAKALKKYLLITGILVWLPIAVTVVWVSIVVSASDQLVNLPLKQWRPQYVL

5	orf98ng-1	MTEPAAEGGKAALKKKYLITGILVWLPIAVTVWVVSIVSASDQLVNLPPKQWRPQYVL	10	20	30	40	50	60
	orf98-1.pep	GFNIPGLGVIVAVLFTGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKKVSESL	70	80	90	100	110	120
10	orf98ng-1	SDSSRSFKTPVLVPFPQSGIWTIAFVSGQVSNVKAALPKDGDYLSVYVPTTNPNTGGYY	130	140	150	160	170	180
	orf98-1.pep	SDSSRSFKTPVLVPFPQSGIWTIAFVSGQVSNVKAALPKDGDYLSVYVPTTNPNTGGYY	130	140	150	160	170	180
15	orf98ng-1	IMVKKSDVRELDMSVDEALKYVISLGMVIPDDLVPKTLAGPMPPEKAELPEQQX	190	200	210	220	230	
	orf98-1.pep	IMVKKSDVRELDMSVDEALKYVISLGMVIPDDLVPKTLAGPMPPEKAELPEQQX	190	200	210	220	230	

Based on this analysis, including the fact that the putative transmembrane domains in the gonococcal protein are identical to the sequences in the meningococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 89

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 749>:

1	ATgAAAACGG	TAGTCTGGAT	TGTCGTCCTG	TTGCCGCCG	CCGTCGGACT
51	GGCGCTGGCT	TCGGGCATTT	ACACCGGCGA	CGTGTATATC	GTACTCGGAC
101	AGACCATGCT	CAGAATCAAC	CTGCACGCCT	TTGTGTTAGG	TTCGCTGATT
151	GCGGTCGTGG	TGTGGTATTT	CTGTTTAAA	TTCAATTATCG	GgGgTACTCA
201	ATATCCCCGA	AAAGATGCAG	CGTTTCGGTT	CGGCnCGTAA	AGGCCkCAAG
251	ssCGsGCTTG	CCTTGAACAA	GGCGGGTTTG	GCGTATTTTG	AAGGGCGCTTT
301	TGAAAAGGCG	GAAGTAGAAG	CCTCACGCGT	GTTGGTCAAC	AAAGtAGGCC
351	GaGAGACAAC	CGGACTTTGG	CATTGATGCT	GrGCGCGCAC	GCCGCCGGAC
401	AGATGGAAAA	CATCGAsTG	CGCGACCGTT	ATCTTGCGGA	AATCGCCAAA
451	CTGCCGGA	AACAGCAGCT	TCCCCGTAT	CTTTGTGTTG	CGGAATCGGC
501	GTTGAACCGG	CGGATTACG	AAGCGGCGGA	AGCCAATCTT	CATGCGGCGG
551	CGAAGATGAA	TGCCAACCTT	ACGCGCCTCG	TGCGTCTGCA	.ATTCTGTAC
601	GCTTTCGACA	GGGGCGACGC	GTGTCAGGTT	CTGGCAAAA	CCGAAAAACT
651	TTCCAAGGCG	GGCGCGTTGG	GCAATCGGA	AATGGAACGG	TATCAAAATT
701	GGGCATATCC	GTCGCCAGCT	GGCGGATGCT	GCCGATGCCG	CCGCTTTGAA
751	AACCTGCCTG	AAGCGGATTC	CCGACAGCCT	CAAAAACGGG	GAATTGAGCG
801	TATCGGTGTC	GGAAAAGTAC	GAACGTTTGG	GACTGTATGC	CGATGCGGTC
851	AAATGGGTCA	AACAGCATTA	TCCGCAsAAC	CGCCGCCCCG	AGCTTTTGGA
901	AGCCTTTGTC	GAAAGCGTGC	GCTTTTGGG	CGAGCGCGAA	CAGCAGAAAG
951	CCATCGATTT	TGCCGATGCT	TGGCTGAAAG	AACAGCCCGA	TAACGCGCTT
1001	CTGCTGATGT	ATCTCGGTGC	GCTCGCCTTC	GGCCGCAAC	TTTGGGGCAA
1051	GGCAAAAGGC	TACCTTGAAG	CGAGCATTGC	ATTAAAGCCG	AGTATTTCCG
1101	CGCGTTTGGT	TCTAACAAAG	GTTTTCGACG	AAATCGGAGA	ACCGCAGAAG
1151	GCGGAGGCGC	AC...			

This corresponds to the amino acid sequence <SEQ ID 750; ORF100>:

1	MKTVVWIVVL	FAAAVGLALA	SGIYTGDIYI	VLGQTMRLIN	LHAFVLGSLI
51	AVVVWYFLFK	FIIGVLNIPE	KMQRFGSARK	GKXXLALNK	AGLAYFEGRF
101	EKAELASRV	LVNKVGRDNR	TLALMLXAHA	AGQMENIXXR	DRYLAETAKL
151	PEKQQLSRYL	LLAESALNRR	DYEAEEANLH	AAAKMNANLT	RLVRLXIRYA
201	FDRGDALQVL	AKTEKLSKAG	ALGXSEMERY	QNWAYRRQLA	DAADAAALKT
251	CLKRIPDSLK	NGELSVSVAE	KYERLGLYAD	AVKWKQHYLP	XNRRPELLEA
301	FVESVRFLGE	REQQKAIDFA	DAWLKEQPDN	ALLMYLGRIL	AFGRKLWGKA

351 KGYLEASIAL KPSISARLVL TKVFDEIGEP QKAEAH...

Further work revealed the complete nucleotide sequence &lt;SEQ ID 751&gt;:

1 ATGAAAACGG TAGTCTGGAT TGTCGTCCTG TTTGCCGCCG CCGTCGGACT  
 51 GCGGCTGGCT TCGGGCATT TACACCGCGA CGTGATATC GACTCGGAC  
 101 AGACCATGCT CAGAATCAAC CTGCACGCTT TTGTGTTAGG TTCGCTGATT  
 151 GCGGTCGTGG TGTGGTATTT CTGTTTAAA TTCATTATCG GCGTACTCAA  
 201 TATCCCCGAA AAGATGCAGC GTTTCGGTTC GCGCGGTAAG GGCCGCAAGG  
 251 CCGCGCTTGC CTTGAACAAG GCGGGTTTGG CGTATTTTGA AGGGCGTTTT  
 301 GAAAAGGCGG AACTAGAAGC CTCACGCGTG TTGGTCAACA AAGAGGCCGG  
 351 AGACAACCGG ACTTTGGCAT TGATGCTGGG CGCGCACGCC GCCGGACAGA  
 401 TGGAAAACAT CGAGCTGCGC GACCGTTATC TTGCGGAAAT CGCCAACTG  
 451 CCGGAAAAC AGCAGCTTTC CCGTTATCTT TTGTGGCGG AATCGGCGTT  
 501 GAACCGGCGC GATTACGAAG CGGCGGAAGC CAATCTTCAT GCGGCGGCGA  
 551 AGATGAATGC CAACCTTACG CGCTCGTGC GTCTGCAACT TCGTTACGCT  
 601 TTCGACAGGG GCGACGCGTT GCAGGTTCTG GCAAAAACCG AAAAATTTTC  
 651 CAAGGCGGGC GCGTTGGGCA AATCGGAAAT GGAACGGTAT CAAAATTTGGG  
 701 CATACCGCCG CCAGCTGGCG GATGCTGCCG ATGCCGCCGC TTTGAAAACC  
 751 TGCCTGAAGC GGATTTCCGA CAGCCTCAA AACGGGGAAT TGAGCGTATC  
 801 GGTTCGCGAA AAGTACGAAC GTTGGGACT GTATGCCGAT GCGGTCAAAT  
 851 GGGTCAACA GCATTATCCG CACAACGCC GCCCGAGCT TTTGGAAGCC  
 901 TTGTTCGAAA CGGTGCGCTT TTTGGGCGAG CGCGAACAGC AGAAAGCAT  
 951 GCATTTTGCC GATGCTTGGC TGAAAGAACA GCCCGATAAC GCGCTTCTGC  
 1001 TGATGTATCT CGGTGCGCTC GCCTACGCC GCAAACTTTG GGGCAAGGCA  
 1051 AAAGGCTACC TTGAAGCGAG CATTGCATTA AAGCCGAGTA TTTCCGCGCG  
 1101 TTTGGTTCTA GCAAAGGTTT TCGACGAAAT CGGAGAACC CAGAAGGCGG  
 1151 AGGCGCAGCG CAACTTGGTT TTGGAAGCCG TCTCCGATGA CGAACGTCAC  
 1201 GCAGCGTTAG AGCAGCATAG CTGA

This corresponds to the amino acid sequence &lt;SEQ ID 752; ORF100-1&gt;:

1 MKTVVWIVVL FAAAVGLALA SGIYTG DVYI VLGQTMLRIN LHAFLVGLSLI  
 51 AVVVWYFLFK FIIGVNIPE KMQRFGSARK GRKAALALNK AGLAYFEGRF  
 101 EKAELEASRV LVNKEAGDNR TLALMLGAHA AGQMENIELR DRYLAEIAKL  
 151 PEKQQLSRYL LLAESALNRR DYEAAEANLH AAKMNANLT RLVRLQLRYA  
 201 FDRGDALQVL AKTEKLSKAG ALGKSEMERY QNWAYRRQLA DAADAAALKT  
 251 CLKRIPDSLK NGELSVSVAE KYERLGLYAD AVKWVKQHYH HNRPELLEA  
 301 FVESVRFLGE REQQKAIDFA DAWLKEQPDN ALLMYLGR LAYGRKLWGKA  
 351 KGYLEASIAL KPSISARLVL AKVFDEIGEP QKAEAQRLV LEAVSDDERH  
 401 AALEQHS\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

40 ORF100 shows 93.5% identity over a 386aa overlap with an ORF (ORF100a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
orf100.pep		MKTVVWIVVLFAAAVGLALAS	SGIYTG DVYI VLGQTMLRIN	LHAFLVGLSLI	AVVVWYFLFK		
orf100a		MKTVVWIVVLFAAAXGLALAS	GIXTG DVYI VLGQTMLRIN	LHAFLVGLSLI	AVVVWYFLFK		
		10	20	30	40	50	60
orf100.pep		FIIGVNIPEKMQRFGSARKG	XXXXLALNKAGLAYFEGRF	EKAELEASRVLVN	KVGRDNR		
orf100a		FIIGVNLXPEKMQRFGSARK	GRKAALALNKAGLAYFEGRF	EKAELEASRVLG	NKEAGDNR		
		70	80	90	100	110	120
orf100.pep		TLALMLXAHAAAGQMENIX	RRDRYLAEIAKLPEKQQLS	RYLLAESALNRRD	YEAAEANLH		
orf100a		TLALMLGAHAAGQMENIEL	RRDRYLAEIAKLPEKQQLS	RYLLAESALNRRD	YEAAEANLH		
		130	140	150	160	170	180
orf100.pep		TLALMLXAHAAAGQMENIX	RRDRYLAEIAKLPEKQQLS	RYLLAESALNRRD	YEAAEANLH		
orf100a		TLALMLGAHAAGQMENIEL	RRDRYLAEIAKLPEKQQLS	RYLLAESALNRRD	YEAAEANLH		
		190	200	210	220	230	240

	orf100.pep	AAAKMNANLTRLVRLXIRYAFDRGDALQVLAKTEKLSKAGALGKSEMERYQNWAYRRQLA
	orf100a	AAAKMNANLTRLVRLQLRYAFDRGDALQVLAKTEKXSKAGAXGKSEMERYQNWAYRRQLX
5		190 200 210 220 230 240
	orf100.pep	DAADAAALKTKLKRIPDSLKNGELSVSVAEKYERLGLYADAVKWVKQHYPKNRRPELLEA
	orf100a	DAADAAALKTKLKRIPDSLKNGELSVSVAEKYERLGLYADAVKWVKQHYPHNRRPELLEA
10		250 260 270 280 290 300
	orf100.pep	FVESVRFLGEREQQKAIDFADAWLKEQPDNALLMYLGLRAFGRLWGKAKGYLEASIAL
	orf100a	FVESVRFLGERDQQKAIDFADAWLKEQPDNALLXYLGLRAYGRKLWGKAKGYLEASIAL
15		310 320 330 340 350 360
	orf100.pep	KPSISARLVLTQVFDEIGEPQKAEAH
	orf100a	KPSISARLVLAQVFDETGEPPQKAEQRNLVLASVAEENRPSAETHX
20		370 380 390 400

The complete length ORF100a nucleotide sequence <SEQ ID 753> is:

25	1	ATGAAAACGG	TAGTCTGGAT	TGTCGTCTCTG	TTTGCCGCCG	CNNTCGGGCT
	51	GGCATTGGCG	TCGGGCATTN	ACACCGGCGA	CGTGTATATC	GTACTCGGAC
	101	AGACCATGCT	CAGAATCAAC	CTGCACGCCT	TTGTGTTAGG	TTTCGTGATT
	151	CCCGTCGTGG	TGTGGTATTT	CCTGTTCAA	TTTCATCATCG	GCGTACTCAA
30	201	TANCCCCGAA	AAGATGCAGC	GTTTCGGTTC	GGCGCGTAAA	GGCCGCAAGG
	251	CCGCGCTTGC	TTTGAAACAAG	GCGGGTTTTGG	CGTATTTTGA	AGGGCGTTTT
	301	GAAAAGGCGG	AACCTGAAGC	CTCGCGCGTA	TTGGGAACA	AAGAGGCGGG
	351	GGATAACCCG	ACTTTGGCAT	TGATGTTGGG	CGCATCATGCC	GCCGGGCAGA
35	401	TGGAAAACAT	CGAGCTGCGC	GACCGTTATC	TTGCGGAAAT	CGCCAAACTG
	451	CCGGAAAAAG	AGCAGCTTTC	CCGTTATCTT	TTGTTGGCGG	AATCGGCGTT
	501	GAACCGGCGC	GATTACGAAG	CGCGGGAAGC	CAATCTTCAT	CGCGCGGCGA
	551	AGATGAATGC	CAACCTTACG	CGCCTCGTGC	GTCTGCAACT	TCGTTACGCT
40	601	TTTCGACAGG	GCGACGCGTT	GCAGGTTCTG	GCAAAAACCG	AAAAANTTTC
	651	CAAGGCGGGC	GCGTNGGGCA	AATCGGAAAT	GGACCGGTAT	CAAAATTGGG
	701	CATACGCGCG	CCAGCTGNCG	GATGCTGCCG	ATGCCCGCCG	TTTGA AAAACC
	751	TGCCTGAAGC	GGATTCCCGA	CAGCCTCAA	AACGGGGAAT	TGAGCGTATC
45	801	GGTTGCGGAA	AAGTACGAAC	GTTTGGGACT	GTATGCCGAT	GCGGTCAAAT
	851	GGGTCAAACA	GCATTATCCG	CACAACCGCC	GACCCGCAACT	TTTGG AAGCN
	901	TTTGTGCGAA	GCGTGCGCTT	TTTGGGCGAA	CGCGATCAGC	AGAAAGCCAT
	951	CGATTTTGCG	GATGCTTGGC	TGAAAGAACA	GCCCGATAAT	GCGCTTCTGC
45	1001	TGANGTATCT	CGGTGCGGTC	GCCTACGGCC	GCAAACTTTG	GGGCAAGGCA
	1051	AAAGGCTACC	TTGAAGCGAG	CATTGCTATTA	AAGCCGAGTA	TTTCCGCGCG
	1101	TTTGGTCTCTG	GCAAAGGTTT	TTGACGAAAC	CGGAGAACCG	CAGAAGGCGG
	1151	AGGCGCAGCG	CAACTTGGTT	TTGGCAAGCG	TTGCCGAGGA	AAACCGNCCT
	1201	TCGCGCGAAA	CCCATTGA			

This encodes a protein having amino acid sequence <SEQ ID 754>:

50	1	<u>MKTVVWIVVL</u>	<u>FAAAXGLALA</u>	SGIXTGDVYI	VLGQTMLRIN	LHAFVLGSLI
	51	<u>AVVVWYFLFK</u>	<u>FIIGVLNXPE</u>	KMQRFGSARK	GRKAALALNK	AGLAYFEGRF
	101	EKAELASRV	LGNKEAGDNR	TLALMLGAHA	AGQMENIELR	DRYLAETAKL
	151	PEKQQLSRV	LLAESALNRR	DYEAEEANLH	AAAKMANLNT	RLRVLQLRYA
55	201	FDRGDALQVL	AKTEKXSKAG	AXGKSEMERY	QNWAYRRQLX	DAADAALKYT
	251	CLKRIPDSLK	NGELSVSVAE	KYERLGLYAD	AVKWKQHYF	HNRRPELLEA
	301	FVESVRLFGE	RDQQKAIDFA	DAWLKEQPDN	ALLLYLGRLL	AYGRKLWGKA
	351	KGYLEASIAL	KPSISARLVL	AKVFDETGEF	QKAEAQORNLY	LASVAEENRP
	401	SAETH*				

ORF100a and ORF100-1 show 95.1% identity in 406 aa overlap:

```

60          10          20          30          40          50          60
orfl00a.pep  MKTVVWIVVLFFAAAXGLALASGIXTGDVYIVLGQTMLRINLHAFVLGSLIAVVVWYFLFK
              |||||
orfl00-1     MKTVVWIVVLFFAAAVGLALASGIYTGDVYIVLGOTMLRINLHAFVLGSLIAVVVWYFLFK

```

		10	20	30	40	50	60
		70	80	90	100	110	120
5	orf100a.pep	FIIGVLNXPEKMQRFGSARKGRKAALALNKAGLAYFEGRFKAELEASRVLGNKEAGDNR					
	orf100-1	FIIGVLNIPEKMQRFGSARKGRKAALALNKAGLAYFEGRFKAELEASRVLVNKEAGDNR					
		70	80	90	100	110	120
10	orf100a.pep	130	140	150	160	170	180
	orf100-1	TLALMLGAHAAGQMENIELRDRLAEIAKLPEKQQLSRYLLLAESALNRRDYEAEEANLH					
		130	140	150	160	170	180
15	orf100a.pep	190	200	210	220	230	240
	orf100-1	AAAKMNANLTRLVRLQLRYAFDRGDALQVLAKTEKXSKAGAXGKSEMERYQNWAYRRQLX					
		190	200	210	220	230	240
20	orf100a.pep	250	260	270	280	290	300
	orf100-1	DAADAAALKTKLRIPDSLKNGELSVSVAEKYERLGLYADAVKWKQHYPHNRRPELLEA					
		250	260	270	280	290	300
25	orf100a.pep	310	320	330	340	350	360
	orf100-1	FVESVRFLGERDQQAIDFADAWLKEQPDNALLXYLGRLAYGRKLWGKAKGYLEASIAL					
		310	320	330	340	350	360
30	orf100a.pep	370	380	390	400		
	orf100-1	KPSISARLVLAKEVDETGEPEQKAEQARNLVLASVAEENRPSA-ETHX					
		370	380	390	400		

Homology with a predicted ORF from *N.gonorrhoeae*

40 ORF100 shows 93.3% identity over a 386 aa overlap with a predicted ORF (ORF100ng) from *N.gonorrhoeae*:

	orf100.pep	MKTVVWIVVLFAAAVGLALASGIYTGDDYIVLGQTMLRINLHAFVLGSLIAVVVWYFLFK	60
	orf100ng	MKTVVWIVVLFAAAVGLALASGIYTGDDYIVLGQTMLRINLHAFVLGSLIAVVVWYFLFK	60
45	orf100.pep	FIIGVLNIPEKMQRFGSARKGKXXLALNKAGLAYFEGRFKAELEASRVLVNKGVDNR	120
	orf100ng	FIIGVLNIPENMRRSGSARKGRKAALALNKAGLAYFEGRFKAELEASRVLGNKEAGDNR	120
50	orf100.pep	TLALMLGAHAAGQMENIXXRDRYLAEIAKLPEKQQLSRYLLLAESALNRRDYEAEEANLH	180
	orf100ng	TLALMLGAHAAGQMENIELRDRLAEIAKLPEKQQLSRYLLLAESALNRRDYEAEEANLH	180
55	orf100.pep	AAAKMNANLTRLVRLXIRYAFDRGDALQVLAKTEKLSKAGALGKSEMERYQNWAYRRQLA	240
	orf100ng	AAAKMNANLTRLVRLQLRYAFDRGDALQVLAKTEKLSKAGALGKSEMERYQNWAYRRQMA	240
	orf100.pep	DAADAAALKTKLRIPDSLKNGELSVSVAEKYERLGLYADAVKWKQHYPHNRRPELLEA	300
60	orf100ng	DAADAAALKTKLRIPDSLKNGELSVSVAEKYERLGLYADAVKWKQHYPHNRRPELLEA	300
	orf100.pep	FVESVRFLGEREQQAIDFADAWLKEQPDNALLMYLGRLAYGRKLWGKAKGYLEASIAL	360
65	orf100ng	FVESVRFLGEREQQAIDFADSWLKEQPDNALLMYLGRLAYGRKLWGKAKGYLEASIAL	360



```
orf100.pep    KPSISARLVLTQVDEIGEPQKAEAH      386
               |||| |||||:||||| :: ||||:
orf100ng     KPSIPARLVLAQVDETAQSQKAEAQRLVLASVAGENRPSAETR 405
```

The complete length ORF100ng nucleotide sequence <SEQ ID 755> is:

5	1	ATGAAACCG	TAGTCTGGAT	TGTTGTCCTG	TTTGCCGCCG	CCGTGCGACT
	51	GGCGCTGGCT	TCGGGCATTT	ACACCGGCGA	CGTGTATATC	GTACTCGGAC
	101	AGACCATGCT	CAGAATCAAC	CTGCACGCCT	TTGTGTTAGG	TTGCGTGATT
	151	CGCGTCGTGG	TGTGGTATTT	CCTGTTTAAA	TTTCATCATC	TCGCTACTCA
	201	TATCCCCGAA	AATATGCGGC	GTTCCGGTTC	GGCGCGGAAA	GGCCGCAAGG
10	251	CCGCGCTTGC	CTTGAATAAG	GCGGGTTTGG	CGTATTTTCA	AGGGCGTTTT
	301	GAAAAGGCGG	AACTCGAAGC	CTCTCGAGTG	TTGGGCAACA	AAGAGGCCGG
	351	AGACAACCGG	ACTTTGGCAT	TGATGCTGGG	CGCGCACGCG	GCAGGACAGA
	401	TGGAAATAT	CGAGCTGCGC	GACCGTTATC	TTGCGGAAAT	CGCCAAACTG
	451	CCGGAAAAAC	AGCAGCTTTC	CCGCTATCTT	CTGCTGGCGG	AATCGGCGTT
15	501	AAACGCCGCG	GATTACGAAG	CGGCGGAAGC	CAATCTTCAT	CGCGCGGCGA
	551	AGATGAATGC	CAACCTTACG	CGCCTCGTGC	GTCTGCAACT	TCGTTACGCG
	601	TTTCGATCGG	GCGATGCGTT	GCAGGTTCTG	GCAAAAaccg	AAAAACTTTT
	651	CAAGGCGGGC	GCGTTGGGCA	AATCGGAAAT	GGAACCGTAT	CAAAATTGGG
	701	CATACCGCCG	CCAGATGGCG	GATGCTGCCG	ATGCCGCCGC	TTTGAAAACC
20	751	TGCCTGAAGC	GGATTCCCGA	CAGCCTCAAA	AACGGGGAAT	TGagcGTATC
	801	GGTTGCGGAA	AAGTAGCAAC	GTTTGGGACT	GTATGCCGAT	CGGTCGAAAT
	851	GGGTCAAACA	GCAATTATCC	CACAACCGCC	GCCCCGAGCT	TTTGGAAGCC
	901	TTTGTGCAAA	GCGTGCGCTT	TTTGGGCGAG	CGCGAACAGC	AGAAAGCCAT
	951	CGATTTTGCC	GATTCTTGCG	TGAAAGAACA	CGCCGATAAC	CGCCTTCTGC
25	1001	TGATGTATCT	CGGCCGGGTC	GCTACGGGCC	GCAAACTTTG	GGGTAAAGGCA
	1051	AAAGGCTACC	TTGAAGCGAG	TATTGCACTG	AAGCCGAGTA	TTCCGGCGCG
	1101	TTTGGTGTGG	GCAAAGGTTT	TTGACGAAAC	CGCACAGTCG	CAAAAAGCCG
	1151	AAGCACAGCG	CAACTTGGTT	TTGGCAAGCG	TTGCCGGGGA	AAACCGCCCT
	1201	TCCGCCGAAA	CCCGTTGA			

30 This encodes a protein having amino acid sequence <SEQ ID 756>:

35

1	MKTVVWIVVL	FAAAVGLALA	SGIYTDVYI	VLGQTMLRIN	LHAFVLGSLI
51	<u>AVVVWYFLFK</u>	<u>FIIGVLNIPE</u>	NMRRSGSARK	GRKAALALNK	AGLAYFEGRF
101	EKAEEASRV	LGNKEAGDNR	TLALMLGAHA	AGQMENIELR	DRYLAEIAKL
151	PEKQQLSRV	LLAESALNRR	DYEAEEANLH	AAAKMNNAL	RLVRLQLRYA
201	FDRGDALQVL	AKTEKLSKAG	ALGKSEMERY	QNWAYRRQMA	DAADAALKT
251	CLKRIPDSLK	NGELSVSVAE	KYERLGLYAD	AVKWWQHYV	HNRRPELLEA
301	FVESVRFLGE	REQQKAIDFA	DSWLKEQPDN	ALLMYLGRL	AYGRKLWGKA
351	KGYLEASIAL	KPSIPARLVL	AKVFDETAQS	QKAEAQRNLV	LASVAGENRP
401	SAETR*				

40 ORF100ng and ORF100-1 show 95.3% identity in 402 aa overlap:

		10	20	30	40	50	60
	orf100-1.pep	MKTVVWIVLFAAAVGLALASGIYTGDDYIVLGQTMLRLNLHAFVLGSLIAVVVWYFLFK					
		:   :					
45	orf100ng	MKTVVWIVLFAAAVGLALASGIYTGDDYIVLGQTMLRLNLHAFVLGSLIAVVVWYFLFK					
		10	20	30	40	50	60
	orf100-1.pep	FIIGVLNIPEKMQRFGSARKGRKAALNALNKAGLAYFEGRFKAELEASRVLVNKEAGDNR					
50		:   :					
	orf100ng	FIIGVLNIPENMRRSGSARKGRKAALNALNKAGLAYFEGRFKAELEASRVLG NKEAGDNR					
		70	80	90	100	110	120
55	orf100-1.pep	TLALMLGAHAAGQMENIELRDRYLAEI AKLP EKQQLSR YLLAESALNRRDY EAAE AN LH					
	orf100ng	TLALMLGAHAAGQMENIELRDRYLAEI AKLP EKQQLSR YLLAESALNRRDY EAAE AN LH					
		130	140	150	160	170	180
60	orf100-1.pep	AAAKMNANLTRLVRLQLRYAFDRGDALQVLAKTEKLSKAGALGKSEMERYQNWAYRRQLA					
	orf100ng	AAAKMNANLTRLVRLQLRYAFDRGDALQVLAKTEKLSKAGALGKSEMERYQNWAYRRQMA					
		190	200	210	220	230	240

		250	260	270	280	290	300
5	orf100-1.pep	DAADAAALKTCLKRI	PDSLKNGELSVSVA	EYERLGLYADAVKWV	KQHYPHNRRPELLEA		
	orf100ng	DAADAAALKTCLKRI	PDSLKNGELSVSVA	EYERLGLYADAVKWV	KQHYPHNRRPELLEA		
		250	260	270	280	290	300
10	orf100-1.pep	FVESVRFLGEREQQA	IDFADAWLKEQPDN	ALLMYLGRLAYGRK	LWGKAKGYLEASIAL		
	orf100ng	FVESVRFLGEREQQA	IDFADSWLKEQPDN	ALLMYLGRLAYGRK	LWGKAKGYLEASIAL		
		310	320	330	340	350	360
15	orf100-1.pep	KPSISARLVLA	KVFDEIGEPQKAE	AQRNLVLEAVSDD	ERHAALQHSX		
	orf100n	KPSIPARLVLA	KVFDETAQSQKAE	AQRNLVLASVAG	ENRPSAETRX		
		370	380	390	400		

Based on this analysis, including the presence of a putative leader sequence, a putative transmembrane domain, and a RGD motif, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 90

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 757>

```

1  ATGATGTTTT  CTGGTTCAA  GCTGTTTCAC  TTGTTTTTTG  TCATTTCGTG
51  GTTTCAGGG  CTGTTTACC  TGCCGAGGAT  TTTCGTCAAT  ATGGCGATGA
101 TTGATGTGCC  GCGCGGCAAT  CCCGAGTATG  TGCGTCTGTC  GGGCATGGCG
151 GTGCGGCTGT  ACCGTTTAT  GTCGCCGTTG  GGCTTCGGCG  CGGTCGTGTT
201 CGGCGCGGCG  ATACCGTTT  CCGCCGGCTG  GTGGGGCAGC  GGCTGGGTAC
251 ACGTCAAAC  GTGTTTGGG  TTGATGCTCT  TGGCTTACCA  GTTGTATTGC
301 GCGTGCTGC  TGCGCCGTT  TCAGGATTAC  AGCAATGCTT  TTTACACCG
351 CTGGTACCG  GTGTTCAAC  AAATCCCCGT  GCTGCTGATG  GTTGCCGCG
401 TGTATsTGGT  CGTGTTCAA  CCGTTTTGA

```

This corresponds to the amino acid sequence <SEQ ID 758; ORF102>:

```

1  MMFSWEKLFH  LFFVISWFAG  LFYLPRI FVN  MAMIDVPRGN  PEYVRLSGMA
51  VRLYRFMSPL  GFGAVVFGAA  IPFAAGWWS  GWVHVKLCLG  LMLLAYQLYC
101 GVLLRRFQDY  SNAFSHRWYR  VFNEIPVLLM  VAALYXVVEK  PF*

```

Further work revealed the complete nucleotide sequence <SEQ ID 759>:

```

1  ATGATGTTTT  CTGGTTCAA  GCTGTTTCAC  TTGTTTTTTG  TCATTTCGTG
51  GTTTCAGGG  CTGTTTACC  TGCCGAGGAT  TTTCGTCAAT  ATGGCGATGA
101 TTGATGTGCC  GCGCGGCAAT  CCCGAGTATG  TGCGTCTGTC  GGGCATGGCG
151 GTGCGGCTGT  ACCGTTTAT  GTCGCCGTTG  GGCTTCGGCG  CGGTCGTGTT
201 CGGCGCGGCG  ATACCGTTT  CCGCCGGCTG  GTGGGGCAGC  GGCTGGGTAC
251 ACGTCAAAC  GTGTTTGGG  TTGATGCTCT  TGGCTTACCA  GTTGTATTGC
301 GCGTGCTGC  TGCGCCGTT  TCAGGATTAC  AGCAATGCTT  TTTACACCG
351 CTGGTACCG  GTGTTCAAC  AAATCCCCGT  GCTGCTGATG  GTTGCCGCG
401 TGTATCTGGT  CGTGTTCAA  CCGTTTTGA

```

This corresponds to the amino acid sequence <SEQ ID 760; ORF102-1>:

```

1  MMFSWEKLFH  LFFVISWFAG  LFYLPRI FVN  MAMIDVPRGN  PEYVRLSGMA
51  VRLYRFMSPL  GFGAVVFGAA  IPFAAGWWS  GWVHVKLCLG  LMLLAYQLYC
101 GVLLRRFQDY  SNAFSHRWYR  VFNEIPVLLM  VAALYLVEFK  PF*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with HP1484 hypothetical integral membrane protein of *H. pylori* (accession number AE000647)

ORF102 and HP1484 show 33% aa identity in 143aa overlap:

```

5      orf102  3  FSWFKLFHLFFVISWFAGLFYLPRI FVN MAMIDVPRGNPEYVRLSGMAVRLYRFMSPLGF 62
      HP1484  8  FLWVKAHFVIAVISWMAALFYLPRLFVYHAENAHKKEFVG VVQIQEK--KLYSFIAS PAM 65

      orf102  63  GAVVFGAAIPFAAG---WWGSGWVHV KLC LGLM L LAYQLYCGVLLRRFQDYSNAFSHRWY 119
      HP1484  66  GFTLITGILMLLIEPTLFKSGGWLHAKLALV VLL LAYHFYCKKCMRELEKDPTRRNARFY 125

      orf102  120 RVFNEIPXXXXXXXXXXXXXFKPF 142
      HP1484  126 RVFNEAP TILMILIVLVVVKPF 148
  
```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF102 shows 99.3% identity over a 142aa overlap with an ORF (ORF102a) from strain A of *N.*

*meningitidis*:

```

20      orf102.pep  MMFSWFKLFHLFFVISWFAGLFYLPRI FVN MAMIDVPRGNPEYVRLSGMAVRLYRFMSPL
      orf102a      MMFSWFKLFHLFFVISWFAGLFYLPRI FVN MAMIDVPRGNPEYVRLSGMAVRLYRFMSPL
                        10      20      30      40      50      60

25      orf102.pep  GFGAVVFGAAIPFAAGW WSGWVHV KLC LGLM L LAYQLYCGVLLRRFQDYSNAFSHRWYR
      orf102a      GFGAVVFGAAIPFAAGW WSGWVHV KLC LGLM L LAYQLYCGVLLRRFQDYSNAFSHRWYR
                        70      80      90      100     110     120

30      orf102.pep  VFNEIPVLLMVAALYXV VFKPFX
      orf102a      VFNEIPVLLMVAALYLVVFKPFX
                        130     140
  
```

The complete length ORF102a nucleotide sequence <SEQ ID 761> is:

```

1  ATGATGTTTT CTTGGTTCAA GCTGTTTCAC TTGTTTTTTG TCATTTCTGTG
51 GTTTGCAGGG CTGTTTACC TCCGAGGAT TTTCGTCAAT ATGGCGATGA
101 TTGATGTGCC GCGCGCAAT CCCGAGTATG TGCGTCTGTC GGGCATGGCG
40 151 GTGCGGCTGT ACCGTTTAT GTCGCCGTTG GGCTTCGGCG CGGTCGTGTT
201 CGGCGCGGCG ATACCGTTTG CCGCCGGCTG GTGGGCGAGC GGCTGGGTAC
251 ACGTCAAAC TGTGTTGGGC TTGATGCTCT TGGCTTACCA GTTGATTGTC
301 GGCGTGCTGC TGCGCCGTTT TCAGGATTAC AGCAATGCTT TTTACACCGC
351 CTGGTACCGC GTGTTCAACG AAATCCCCGT GCTGCTGATG GTTGCCGCGC
45 401 TGTATCTGGT CGTGTTCAAA CCGTTTTGA
  
```

This encodes a protein having amino acid sequence <SEQ ID 762>:

```

1  MMFSWFKLFH LFFVISWFA G LFYLPRI FVN MAMIDVPRGN PEYVRLSGMA
51 VRLYRFMSPL GFGAVVFGAA IPFAAGW WSG GWVHV KLC LG LMLLAYQLY C
101 GVLLRRFQDY SNAFSHRWYR VFNEIPVLLM VAALYLVVFK PF*
  
```

ORF102a and ORF102-1 show complete identity in 142 aa overlap:

```

55      orf102a.pep  MMFSWFKLFHLFFVISWFAGLFYLPRI FVN MAMIDVPRGNPEYVRLSGMAVRLYRFMSPL
      orf102-1      MMFSWFKLFHLFFVISWFAGLFYLPRI FVN MAMIDVPRGNPEYVRLSGMAVRLYRFMSPL
                        10      20      30      40      50      60
  
```

		70	80	90	100	110	120
	orf102a.pep	GFGAVVFGAAIPFAAGWWSG	WVHVVKLCLGLMLLAYQLYC	GVLLRRFQDYSNAF	SHRWYR		
5	orf102-1	GFGAVVFGAAIPFAAGWWSG	WVHVVKLCLGLMLLAYQLYC	GVLLRRFQDYSNAF	SHRWYR		
		70	80	90	100	110	120
10	orf102a.pep	VFNEIPVLLMVAALYL	VVFKPFX				
	orf102-1	VFNEIPVLLMVAALYL	VVFKPFX				
		130	140				

### Homology with a predicted ORF from *N.gonorrhoeae*

15 ORF102 shows 97.9% identity over a 142 aa overlap with a predicted ORF (ORF102ng) from *N. gonorrhoeae*:

	orf102.pep	MMFSWFKLFLHFFVISWFAGLFYLPRI	FVNMAMIDVPRGNPEYVRLSGMAVRLYR	FMSPL	60
	orf102ng	MMFSWFKLFLHFFVISWFAGLFYLPRI	FVNMAMIDAPRGNPEYVRLSGMAVRLYR	FMSPL	60
20	orf102.pep	GFGAVVFGAAIPFAAGWWSG	WVHVVKLCLGLMLLAYQLYC	GVLLRRFQDYSNAF	SHRWYR 120
	orf102ng	GFGAVVFGAAIPFAAGWWSG	WVHVVKLCLGLMLLAYQLYC	GVLLRRFQDYSNAF	SHRWYR 120
25	orf102.pep	VFNEIPVLLMVAALYX	VVFKPF	142	
	orf102ng	VFNEIPVLLMVAALYL	VVFKPF	142	

The complete length ORF102ng nucleotide sequence <SEQ ID 763> is:

30	1	ATGATGTTTT	CTGGTTCAA	GCTGTTTCAC	TTGTTTTTTG	TCATTTCGTG
	51	GTTTGCAGGG	CTGTTTTACC	TGCCGAGGAT	TTTCGTCAAT	ATGGCGATGA
	101	TTGATGCGCC	CGCGGCAAT	CCCAGATATG	TGCGCCTGTC	GGGATGGCG
	151	GTGCGGTGT	ACCGTTTTAT	GTCGCCTTTG	GGTTTCGGCG	CGGTCGTGTT
	201	CGGCGCGCG	ATACCGTTTG	CCGCgggccc	GTGGGCGagc	ggctggGTTT
	251	ACGTCAAAC	GTGTTTGGGC	TTGATGCTCT	TGGCTTATCA	GTTGTATTGC
35	301	GGCGTGCTGC	TGCGCCGTTT	TCAGGATTAC	AGCAATGCTT	TTTCACACCG
	351	CTGGTACCGC	GTGTTCAAag	aAATCCCCGT	GCTGCTGATG	GTTGCCGCGC
	401	TGTATCTGGT	CGTGTTCAA	CCGTTTTGA		

This encodes a protein having amino acid sequence <SEQ ID 764>:

40	1	MMFSWFKLFH	LFFVISWFAG	LFYLPRI	FVNMAMIDAPRGN	PEYVRLSGMA
	51	VRLYRFSPL	GFGAVVFGAA	IPFAAGRWS	GWVHVVKLCLG	LMLLAYQLYC
	101	GVLLRRFQDY	SNAFSHRWYR	VFNEIPVLLM	VAALYL	VVFKPF*

ORF102ng and ORF102-1 show 98.6% identity in 142 aa overlap:

		10	20	30	40	50	60
45	orf102-1.pep	MMFSWFKLFLHFFVISWFAGLFYLPRI	FVNMAMIDVPRGNPEYVRLSGMAVRLYR	FMSPL			
	orf102ng	MMFSWFKLFLHFFVISWFAGLFYLPRI	FVNMAMIDAPRGNPEYVRLSGMAVRLYR	FMSPL			
		10	20	30	40	50	60
50	orf102-1.pep	GFGAVVFGAAIPFAAGWWSG	WVHVVKLCLGLMLLAYQLYC	GVLLRRFQDYSNAF	SHRWYR		
	orf102ng	GFGAVVFGAAIPFAAGRWSG	WVHVVKLCLGLMLLAYQLYC	GVLLRRFQDYSNAF	SHRWYR		
		70	80	90	100	110	120
55	orf102-1.pep	VFNEIPVLLMVAALYL	VVFKPFX				
	orf102ng	VFNEIPVLLMVAALYL	VVFKPFX				
		130	140				

In addition, ORF102ng shows significant homology to a membrane protein from *H. pylori*:

gi|2314656 (AE000647) conserved hypothetical integral membrane protein  
[*Helicobacter pylori*] Length = 148  
Score = 79.2 bits (192), Expect = 1e-14  
Identities = 50/147 (34%), Positives = 68/147 (46%), Gaps = 13/147 (8%)

5 Query: 3 FSWFKLFHLFFVISWVAGLFYLPRIFFVNMAMIDAPRGNPEYVRLSGMAVRLYRFMSPLGF 62  
F W K FH+ VISW A LFYLP+V A + V++ +LY F++  
10 Sbjct: 8 FLWVKAFHVIAVISWMAALFYLPRLFYVHAENAHKKEFVGVVQIQEK--KLYSFIASPM 65

Query: 63 GAVVFGAAIP-----FAAGRWSGWVHVKLCLGLMLLAYQLYCGVLLRRFQDYSNAFS 115  
G + + F +G GW+H KL L ++LLAY YC +R + +  
15 Sbjct: 66 GFTLITGILMLLIEPTLFKSG---GWLHAKLALVLLLAYHFYCKKCMRELEKDPTRRN 121

Query: 116 HRWYRVFNEIPXXXXXXXXXXXXXFKPF 142  
R+YRVFNE P KPF  
15 Sbjct: 122 ARFYRVFNEAPTILMILIVILVVVKPF 148

Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

## 20 Example 91

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 765>:

1 ATGGCAAAAA TGATGAAATG GCGGGCTGTT GCGGCGGTCG CCGCGGCAGC  
51 GGTTTGGGGC GGATGGTCTT AACTGAAGCC CGAGCCGCAC GTGCTTGATA  
101 TTACGGAAAC GGTCAGGCGC GGC // ....  
25 //.. ATTTCTGTTA CGATTTTGTC CGAACCGGAT ACGCGGATTA AGGCGAAGCT  
51 CGACAGCGTC GACCCCGGGC TGACCACGAT GTCGTCGGGC GGTACAACA  
101 GCAGTACGGA TACGGCTTCC AATGCGGTCT ACTATTATGC CCGTTCGTTT  
151 GTGCCGAATC CGGACGGCAA ACTCGCCACG GGGATGACGA CGCAGAATAC  
201 GGTTGAAATC GACGGCGTGA AAAATGTGCT GATTATTCCG TCGCTGACCG  
301 TGAAAAATCG CCGCGGCAAG GCGTTTGTG GCGTGTGGG TCGGACGGC  
301 AAGGCGGCGG AACCGGAAAT CCGGACCGGT ATGAGAGACA GTATGAATAC  
351 CGAAGTAAAA AGCGGGTGA AAGAGGGGGA CAAAGTGATC ATCTCCGAAA  
401 TAACCGCCGC CGAGCAACAG GAAAGCGGCG AACGCGCCCT AGGCGGCCCG  
451 CCGCGCGGAT AA

35 This corresponds to the amino acid sequence <SEQ ID 766; ORF85>:

1 MAKMMKWAIV AAVAAAAVWG GWS.LKPEPH VLDITETVRR G.....  
51 .....  
101 .....  
151 .....  
40 201 .....I SFTILSEPT  
251 PIKAKLDSVD PGLTTMSSGG YNSSTDASN AVYYYARFV PNPDKLATG  
301 MTTQNTVEID GVKNVLIIPS LTVKNRGGKA FVRVLGADGK AAEREIRTGM  
351 RDSMNTVEVKS GLKEGDKVVI SEITAAEQQE SGERALGGPP RR\*

Further work revealed the further partial nucleotide sequence <SEQ ID 767>:

45 1 ..GTATCGGTCG GCGCGCAGGC ATCGGGGCAG ATTAAGATAC TTTATGTCAA  
51 ACTCGGGCAA CAGGTTAAAA AGGGCGATTG GATTGCGGAA ATCAATTCTGA  
101 CCTCGCAGAC CAATACGCTC AATACGGAAA AATCCAAGTT GGAAACGTAT  
151 CAGGCGAAGC TGGTGTGCGC ACAGATTGCA TTGGGCAGCG CGGAGAAGAA  
201 ATATAAGCGT CAGGCGGCGT TATGGAAGGA AAACGCGACT TCCAAAGAGG  
50 251 ATTTGGAAG CCGCAGGAT GCGTTTGCCG CCGCAAAGC CAATGTTGCC  
301 GAGCTGAAG CTTTAATCAG ACAGAGCAAA ATTTCCATCA ATACCGCCGA  
351 GTCGGAATTG GGCTACACGC GCATTACCGC AACGATGGAC GGCACGGTGG  
401 TGCGGATTCT CTGGAAGAG GGCAGACTG TGAACGCGGC GCAGTCTACG  
451 CCGACGATTG TCCAATTGGC GAATCTGGAT ATGATGTTGA ACAAAATGCA  
55 501 GATTGCCGAG GCGGATATTA CCAAGGTGAA GGCGGGGCAG GATATTTCGT  
551 TTACGATTTT GTCCGAACCG GATACGCCGA TTAAGGCGAA GCTCGACAGC  
601 GTCGACCCCG GGCTGACCAC GATGTCGTCG GCGGGTTACA ACAGCAGTAC

5  
651 GGATACGGCT TCCAATGCGG TCTACTATTA TGCCCGTTTCG TTTGTGCCGA  
701 ATCCGGACGG CAAACTCGCC ACGGGGATGA CGACGCAGAA TACGGTTGAA  
751 ATCGACGGCG TGAAAAATGT GCTGATTATT CCGTCGCTGA CCGTGAAAAA  
801 TCGCGGCGGC AAGGCGTTTG TCGCGGTGT GGGTGCGGAC GGCAAGGCGG  
851 CGGAACGCGA AATCCGGACC GGTATGAGAG ACAGTATGAA TACCGAAGTA  
901 AAAAGCGGGT TGAAAGAGGG GGACAAAGTG GTCATCTCCG AAATAACCGC  
951 CGCCGAGCAA CAGGAAAGCG GCGAACGCGC CCTAGGCGGC CCGCCGCGCC  
1001 GATAA

This corresponds to the amino acid sequence <SEQ ID 768; ORF85-1>:

10  
1 ..VSVGAQASGQ IKILYVKLGQ QVKKGLIAE INSTSQNTL NTEKSKLETY  
51 QAKLVSAQIA LGSAEKKYKR QAALWKENAT SKEDLESAQD AFAAAKANVA  
101 ELKALIRQSK ISINTAESEL GYTRITATMD GTVVAILVEE GQTVNAAQST  
151 PTIVQLANLD MMLNKMQUIAE GDITKVKAGQ DISFTILSEP DTPIKAKLDS  
201 VDPGLTMS GGYNSSTD TA SNAVYYARS FVPNPDGKLA TGMTTQNTVE  
15  
251 IDGVKNVLI PSLTVKNRGG KAFVRVLGAD GKAAEREIRT GMRDSMNTEV  
301 KSGLKEGDKV VISEITAAEQ QESGERALGG PPRR\*

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF85 shows 87.8% identity over a 41aa overlap and 99.3% identity over a 153aa overlap with an ORF (ORF85a) from strain A of *N. meningitidis*:

20  
orf85.pep                   10           20           30           40  
                  MAKMMKWAAVA AAAAAVWGGWS-LKPEPHVLDITETVRRG  
                  |||||  
25 orf85a                   10           20           30           40           50           60  
                  MAKMMKWAAVA AAAAAVWGGWSY LKPEPQAAYITETVRRGDISRTVSATGEISPSNLVS  
                  |||||  
                  //  
                  80           90           100  
orf85.pep                   .....ISFTILSEPDTPIKAKLDSVDPGLTMS  
                  |||||  
30 orf85a                   TIVQLANLDMMLNKMQUIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPGLTMS  
210           220           230           240           250           260  
orf85.pep                   110           120           130           140           150           160  
                  GYNSSTD TASNAVYYARSFV PNPDKLATGMTTQNTVEIDGVKNVLIIPSLTVKNRGGK  
                  |||||  
35 orf85a                   GYNSSTD TASNAVYYARSFV PNPDKLATGMTTQNTVEIDGVKNVLIIPSLTVKNRGR  
270           280           290           300           310           320  
orf85.pep                   170           180           190           200           210           220  
                  AFVRVLGADGKAAEREIRTGM RDSMNTEVKSGLKEGDKV VISEITAAEQQESGERALGGP  
                  |||||  
40 orf85a                   AFVRVLGADGKAAEREIRTGM RDSMNTEVKSGLKEGDKV VISEITAAEQQESGERALGGP  
330           340           350           360           370           380  
orf85.pep                   230  
                  PRRX  
                  ||||  
45 orf85a                   PRRX  
                  390

50 The complete length ORF85a nucleotide sequence <SEQ ID 769> is:

1 ATGGCAAAAA TGATGAAATG GCGGCTGTT GCGGCGGTCG CCGCGGCAGC  
51 GGTGTGGGGC GGATGGTCTT ATCTGAAGCC CGAGCCGCAG GCTGCTTATA  
101 TTACGGAAAC GGTGAGGCGC GCGGACATCA GCCGGACGGT TTCTGCAACA  
151 GGGGAGATTT CGCCGTCCAA CCTGGTATCG GTCGCGCGCG AGGCATCGGG  
55 201 GCAGATTAAG AAATTTATG TCAAACCTCG GCAACAGGTT AAAAAGGGCG  
251 ATTTGATTGC GGAAATCAAT TCGACCTCGC AGACCAATAC GCTCAATACG  
301 GAAAAATCCA AATTGGAAC GTATCAGGCG AAGCTGGTGT CCGCACAGAT  
351 TGCATTGGGC AGCGCGGAGA AGAAATATAA GCGTCAGGCG GCGTTGTGGA  
401 AGGATGATGC GACCGCTAAA GAAGATTTGG AAAGCGCACA GGATGCGCTT  
60 451 GCCCGCGCCA AAGCCAATGT TGCCGAGCTG AAGGCTCTAA TCAGACAGAG

501 CAAAATTTCC ATCAATACCG CCGAGTCGGA ATTGGGCTAC ACGCGCATT  
 551 CCGCAACGAT GGACGGCAGC GTGGTGGCGA TTCTCGTGGA AGAGGGGCAG  
 601 ACTGTGAACG CCGCGCAGTC TACGCCGACG ATGTGCCAAT TGGCGAATCT  
 651 GGATATGATG TTGAACAAAA TGCAGATTGC CGAGGGCGAT ATTACCAAGG  
 701 TGAAGGCGGG GCAGGATATT TCGTTTACGA TTTGTGCCGA ACCGGATACG  
 751 CCGATTAAGG CGAAGCTCGA CAGCGTCGAC CCCGGGCTGA CCACGATGTC  
 801 GTCGGGCGGC TACAACAGCA GTACGGATAC GGCTTCCAAT GCGGTCTACT  
 851 ATTATGCCCG TTCGTTTGTG CCGAATCCGG ACGGCAAACG CGCCACGGGG  
 901 ATGACGACGC AGAATACGGT TGAAATCGAC GGTGTGAAAA ATGTGCTGAT  
 951 TATTCCGTCG CTGACCGTGA AAAATCGCGG CGGCAGGGCG TTTGTGCGCG  
 1001 TGTGGGTGC AGACGGCAAG GCGGCGGAAC GCGAAATCCG GACCGGTATG  
 1051 AGAGACAGTA TGAATACCGA AGTAAAAAGC GGGTTGAAAG AGGGGGACAA  
 1101 AGTGGTCATC TCCGAAATAA CCGCCGCCGA GCAGCAGGAA AGCGGCGAAC  
 1151 GCGCCCTAGG CGGCCCGCCG CGCCGATAA

15 This encodes a protein having amino acid sequence <SEQ ID 770>:

1 MAKMMKWAIV AAVAAAAVWG GWSYLKPEPQ AAYITETVRR GDISRTVSAT  
 51 GEISPSNLVS VQAQASGQIK KLYVKLGQV KKGDLIAEIN STSQNTLNLT  
 101 ESKKLETYQA KLVSAQIALG SAEKKYKRQA ALWKDDATAK EDLESAQDAL  
 151 AAKANVAEL KALIRQSKIS INTAESELGY TRITATMDGT VVAILVEEGQ  
 201 TVNAAQSTPT IVQLANLDM LNMQIAEGD ITKVKAGQDI SFTILSEPDT  
 251 PIKAKLDSVD PGLTTMSSGG YNSSTDASN AVYYYARFV PNPDKLATG  
 301 MTTQNTVEID GVKNVLIIPS LTVKNRGGRA FVRVLGADGK AAEREIRTGM  
 351 RDSMNTEVKS GLKEGDKVVI SEITAAEQQE SGERALGGPP RR\*

ORF85a and ORF85-1 show 98.2% identity in 334 aa overlap:

25 orf85a.pep 30 40 50 60 70 80  
 PQAAYITETVRRGDISRTVSATGEISPSNLVSVGAQASGQIKKLYVKLGQVKKGDLIAE  
 orf85-1 VSVGAQASGQIKILYVKLGQVKKGDLIAE  
 30 10 20 30  
 30 orf85a.pep 90 100 110 120 130 140  
 INSTSQNTLNTEKSKLETYQAKLVSAQIALGSAEKKYKRQAALWKDDATAKEDLESAQD  
 orf85-1 INSTSQNTLNTEKSKLETYQAKLVSAQIALGSAEKKYKRQAALWKENATSKEDLESAQD  
 35 40 50 60 70 80 90  
 40 orf85a.pep 150 160 170 180 190 200  
 ALAAAKANVAELKALIRQSKISINTAESELGYTRITATMDGTVVAILVEEGQTVNAAQST  
 orf85-1 AFAAAKANVAELKALIRQSKISINTAESELGYTRITATMDGTVVAILVEEGQTVNAAQST  
 100 110 120 130 140 150  
 45 orf85a.pep 210 220 230 240 250 260  
 PTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPLTTMSS  
 orf85-1 PTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPLTTMSS  
 160 170 180 190 200 210  
 50 orf85a.pep 270 280 290 300 310 320  
 GGYNSSTDASNNAVYYYARFVNPDKLATGMTTQNTVEIDGVKNVLIIPSLTVKNRGG  
 orf85-1 GGYNSSTDASNNAVYYYARFVNPDKLATGMTTQNTVEIDGVKNVLIIPSLTVKNRGG  
 220 230 240 250 260 270  
 55 orf85a.pep 330 340 350 360 370 380  
 RAFVRVLGADGKAAEREIRTGM RDSMNTEVKSGLKEGDKVVI SEITAAEQQESGERALGG  
 orf85-1 KAFVRVLGADGKAAEREIRTGM RDSMNTEVKSGLKEGDKVVI SEITAAEQQESGERALGG  
 280 290 300 310 320 330  
 60 orf85a.pep 390  
 PPRRX  
 orf85-1 PPRRX

65 Figure 19D shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF85a..

Homology with a predicted ORF from *N.gonorrhoeae*

ORF85 shows a high degree of identity with a predicted ORF (ORF85ng) from *N.gonorrhoeae*:

```

ORF85      1 MAKMMKWA AVAAVAAA VGGWS.LKPEPHVLDITETVRRG..... 40
|||||
5  ORF85ng  1 MAKMMKWA AVAAVAAA VGGWSY LKPEPQAAYITEAVRRGDISRTVSAT 50

ORF85      .....ISFTILSEPDT 250
|||||
10 ORF85ng  201 TVNAAQSTPTIVQLANLDMMLNKMQIAEGDITKV KAGQDISFTILSEPDT 250

ORF85      251 PIKAKLDSVD PGLTTMSSGGYNSSTDASNAVYYYARSFV PNPDPGKLATG 300
|||||
15 ORF85ng  251 PIKAKLDSVD PGLTTMSSGGYNSSTDASNAVYYYARSFV PNPDPGKLATG 300

ORF85      301 MTTQNTVEIDGVKNVLLIPSLTVKNRGGKAFVRVLGADGKAAEREIRTGM 350
|||||
ORF85ng    301 MTTQNTVEIDGVKNVLLIPSLTVKNRGGKAFVRVLGADGKAVEREIRTGM 350

20 ORF85      152 RDSMNTEVKSGLKEGDKVVI SEITAAEQQESGERALGGPPRR 393
:|||||
ORF85ng    351 KDSMNTEVKSGLKEGDKVVI SEITAAEQQESGERALGGPPRR 393

```

The complete length ORF85ng nucleotide sequence <SEQ ID 771> is:

```

25 1 ATGGCAAAAA TGATGAAATG GCGCGCTGTT GCGGCGGTCG CCGCGGCaac
51 GGT TTGGGGC GGATGGTCTT ATCTGAAGCC CGAACCGCAG GCTGCTTATA
101 TTACGGAaac ggTCAGGCGC GCGCATATCA GCCGACGGT TTCCGCGACG
151 GgcgAGATTT CGCCGTCCAA CCTGGTATCG GTCGGCGCGC AGGCTTCGGG
201 GCAGATTAaa AAGCTTTATG TCAAACTCGG GCAACAGGTC AAAAAGGGCG
30 251 ATTTGATTGC GGAAATCAAT TCGACCACGC AGACCAACAC GATCGATATG
301 GAAAAATCCA AATTGGAaAC GTATCAGGCG AAGCTGGTGT CCGCACAGAT
351 TGCATTGGGC AGCGCGGAGA AGAAATATAA GCGTCAGGCG GCGTTGTGGA
401 AGGATGATGC GACCTCTAAA GAAGATTGGA AAAGCGCGCA GGATGCGCTT
451 GCCGCCGCCA AAGCCAATGT TGCCGAGTTG AAGGCTTTAA TCAGACAGAG
501 CAAAATTTCC ATCAATACCG CCGAGTCGGA TTTGGGCTAC ACGCGCATTa
35 551 CCGCGACGAT GGACGGCAGC GTGGTGGCGA TTCCCGTGGA AGAGGGGCAG
601 ACTGTGAACG CGGCGCAGTC TACGCCGACG ATTGTCCAAT TGGCGAATCT
651 GGATATGATG TTGAACAAAA TGCAGATTGC CGAGGGCGAT ATTACCAAGG
701 TGAAGCGGGG GCAGGATATT TCGTTTACGA TTTGTGCCGA ACCGGATACG
40 751 CCGATTAAGG CGAAGCTCGA CAGCGTCGAC CCCGGGCTGA CCACGATGTC
801 GTCGGGCGGC TACAACAGCA GTACGGATAC GGCTTCCAAT GCGGTCTATT
851 ATTATGCCCG TTCGTTTGTG CCGAATCCGG ACGGCAAACT CGCCACGGGG
901 ATGACGACGC AGAATACGGT TGAATCGAC GGTGTGAAAA ATGTGTTGCT
951 TATTCCGTCG CTGACCGTGA AAAATCGCGG CGGCAAGGCG TTCGTACGCG
45 1001 TGTTGGGTGC GGACGGCAAG GCAGTGGAAC GCGAAATCCG GACCGGTATG
1051 AAAGACAGTA TGAATACCGA AGTGAaaAG GGGTTGaaAG AGGGGGACAA
1101 AGTGGTCATC TCCGAAATAA CCGCCGCCGA GCAGCAGGAA AGCGGCGAAC
1151 GCGCCCTAGG CCGCCCGCG CGCCGATAA

```

This encodes a protein having amino acid sequence <SEQ ID 772>:

```

50 1 MAKMMKWA AVAAVAAA VWG GWSY LKPEPQ AAYITEAVRR GDISRTVSAT
51 GEISPSNLVS VGAQASGQIK KLYVKLGQV KKGDLIAEIN STTQNTIDM
101 EKSKLETYQA KLVSAQIALG SAEKKYKRQA ALWKDDATSK EDLESAQDAL
151 AAAKANVAEL KALIRQSKIS INTAESDLGY TRITATMDGT VVAIPVEEGQ
201 TVNAAQSTPT IVQLANLDM LNMQIAEGD ITKV KAGQDI SFTILSEPDT
55 251 PIKAKLDSVD PGLTTMSSGG YNSSTDASN AVYYYARFV PNPDPGKLATG
301 MTTQNTVEID GVKNVLLIPS LTVKNRGGKA FVRVLGADGK AVEREIRTGM
351 KDSMNTEVKS GLKEGDKVVI SEITAAEQQE SGERALGGPP RR*

```

ORF85ng and ORF85-1 show 96.1% identity in 334 aa overlap:

```

60 orf85ng      30          40          50          60          70          80
PQAAYITETVRRGDISRTVSATGEISPSNLVSVGAQASGQIKKLYVKLGQQVKKGDLIAE
orf85-1      |||||||
VSVGAQASGQIKILYVKLGQQVKKGDLIAE

```



				10	20	30		
		90	100	110	120	130	140	
5	orf85ng	INSTTQTNTIDMEKSKLETYQAKLVSAQIALGSAEKYKQQAALWKDDATSKEDLESAQD						
	orf85-1	INSTSQNTNTLNTTEKSKLETYQAKLVSAQIALGSAEKYKQQAALWKENATSKEDLESAQD						
		40	50	60	70	80	90	
		150	160	170	180	190	200	
10	orf85ng	ALAAAKANVAELKALIRQSKISINTAESDLGYTRITATMDGTVVAIPVEEGQTVNAAQST						
	orf85-1	AFAAAKANVAELKALIRQSKISINTAESELGYTRITATMDGTVVAILVEEGQTVNAAQST						
		100	110	120	130	140	150	
		210	220	230	240	250	260	
15	orf85ng	PTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPGLTTMSS						
	orf85-1	PTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPGLTTMSS						
		160	170	180	190	200	210	
		270	280	290	300	310	320	
20	orf85ng	GGYNSSTD TASNAVYYYARSFVNPDPGKLATGMTTQNTVEIDGVKNVLLIPSLTVKNRGG						
	orf85-1	GGYNSSTD TASNAVYYYARSFVNPDPGKLATGMTTQNTVEIDGVKNVLIIPSLTVKNRGG						
25		220	230	240	250	260	270	
		330	340	350	360	370	380	
	orf85ng	KAFVRVLGADGKAVEREIRTGMKDSMNTVEVKSGLKEGDKVVI SEITAAEQQESGERALGG						
30	orf85-1	KAFVRVLGADGKAAEREIRTGMRDSMNTVEVKSGLKEGDKVVI SEITAAEQQESGERALGG						
		280	290	300	310	320	330	
		390						
	orf85ng	PPRRX						
35								
	orf85-1	PPRRX						

In addition, ORF85ng shows significant homology to an *E.coli* membrane fusion protein:

gi|1787104 (AE000189) o380; 27% identical (27 gaps) to 332 residues from  
membrane fusion protein precursor, MTRC\_NEIGO SW: P43505 (412 aa) [Escherichia  
coli] Length = 380  
Score = 193 bits (485), Expect = 2e-48  
Identities = 120/345 (34%), Positives = 182/345 (51%), Gaps = 13/345 (3%)

Query: 29 PQAAITETVRRGDISRTVSATGEISPSNLVSVGAQASGQIKKLYVKLGQVKKGDIAE 88  
P Y T VR GD+ ++V ATG++ V VGAQ SGQ+K L V +G +VKK L+  
Sbjct: 41 PVPTYQTLIVRPGLDQQSVLATGKLDALRKVDVGAQVSGQLKTL SVAIGDKVKKDQLLG V 100

Query: 89 INSTTQNTIDMEKSKLETYQAKLVSAQIALGSAEKYKQQAALWKDDATSKEXXXXXXX 148  
I+ N I ++ L +A+ A+ L A Y RQ L + A S++  
Sbjct: 101 IDPEQAENQIKEVEATLMELRAQRQQAELKLARVTYSRQORLAQTKAVSQQDLDTAAT 160

Query: 149 XXXXXXXXXXXXXXXXIRQSKISINTAESDLGYTRITATMDGTVVAIPVEEGQTVNAAQST 208  
I++++ S++TA+++L YTRI A M G V I +GQTV AAQ  
Sbjct: 161 EMAVKQAQIGTIDAQIKRNQASLDTAKTNLDYTRIVAPMAGEVTQITTLQGQTVIAAQA 220

Query: 209 PTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPGLTTMSS 268  
P I+ LA++ ML K Q++E D+ +K GQ FT+L +P T + ++ V P  
Sbjct: 221 PNILTLADMSAMLVKAQVSEADVHLKPGQKAWFTVLGDP LTRYEGQIKDVL P----- 273

Query: 269 GGYNSSTD TASNAVYYYARSFVNPDPGK LATGMTTQNTVEIDGVKNVLLIPSLTVKNRGG 328  
+ + ++A++YYAR V PNP+G L MT Q +++ VKNVL IP + + G  
Sbjct: 274 -----TPEKVNDAIFYARFEV PNPENGLRLDMTAQVHIQLTDVKNVLTIPLSALGDPVG 328

Query: 329 KAFVRV-LGADGKAVEREIRTGMKDSMNTVEVKSGLKEGDKVVI SE 372  
+V L +G+ ERE+ G ++ + E+ GL+ GD+VVI E  
Sbjct: 329 DNRYKVKLLRNGETREREVTIGARNDDVEIVKGLEAGDEVVIGE 373

Based on this analysis, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF85-1 (40.4kDa) was cloned in the pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 19A shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 19B), FACS analysis (Figure 19C), and ELISA (positive result). These experiments confirm that ORF85-1 is a surface-exposed protein, and that it is a useful immunogen.

### Example 92

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 773>:

```

1  ..ATTCCCGCCA CGATGACATT TGAACGCAGC GGCAATGCTT ACAAATCGT
51  TTCGACGATT AAAGTGCCGC TATACAATAT CCGTTTCGAG TCCGGCGGTA
101 CGGTTGTTCG CAATACCCTG CACCCTACCT ACTATAGAGA CATACGCAGG
151 GGCAAACTGT ATGCGGAAgc CAAATTCGCC GACgGcAGCG TAACTTACGG
201 CAAAGCGGGC GAGAGCAAAA CCGAGCAAAG CCCCAGGCT ATGGATTGTG
251 TCACGCTTGC CTGGCAGTTG GCGGCAAATG ACGCGAAACT CCCCCGGGG
301 CTGAAATCA CCAACGCAA AAAACTTTAT TCCGTCGGCG GTTTGAATAA
351 GGCGGGTACA GAAAAATACA GCATAGGCGG CGTGGAACC GAATCGTCA
401 AATATCGGT GCGGCGCGG GACGATGCGG TAATGTATT cTTCGACCG
20  451 TCCCTGAACA ATATTCCGGC ACAAATCGGC TATACGACG ACGGCAAAAC
501 CTATACGCTG AAACCAAAT CCGTGCAGAT CAACGGCCAG GCAGCCAAAC
551 CGTAA

```

This corresponds to the amino acid sequence <SEQ ID 774; ORF120>:

```

1  ..IPATMTFERS GNAYKIVSTI KVPLYNIRFE SGGTVVGNTL HPTYYDIRR
25  51  GKLYAEAKFA DGSVTYGKAG ESKTEQSPKA MDLFTLAWQL AANDAKLPPG
101  LKITNGKKLY SVGGLNKAGT GKYSIGGVET EVVKYRVRRG DDAVMYFFAP
151  SLNNIPAQIG YTDDGKTYTL KLKSVQINGQ AAKP*

```

Further work revealed the complete nucleotide sequence <SEQ ID 775>:

```

1  ATGATGAAGA CTTTAAAAA TATATTTTCC GCGGCCATT TGTCCGCGC
30  51  CCTGCCGTGC GCGTATGCGG CAGGGCTGCC CCAATCCGCC GTGCTGCACT
101  ATTCCGGCAG CTACGGCATT CCCGCCACGA TGACATTGA ACGCAGCGGC
151  AATGCTTACA AAATCGTTTC GACGATTAAA GTGCCGTAT ACAATATCCG
201  TTTCGAGTCC GCGGTACGG TTGTGCGCAA TACCCTGCAC CCTACCTACT
35  251  ATAGAGACAT ACGCAGGGGC AAATGTATG CGGAAGCCAA ATTCGCCGAC
301  GGCAGCGTAA CTTACGGCAA AGCGGGCGAG AGCAAAACCG AGCAAAAGCCC
351  CAAGGCTATG GATTGTTC CGCTTGCTG GCAGTTGGCG GCAAATGACG
401  CGAAACTCCC CCCGGGGCTG AAAATACCA ACGGCAAAA ACTTTATTCC
451  GTCGGCGGTT TGAATAAGGC GGGTACAGGA AAATACAGCA TAGCGGCGT
501  GGAAACCGAA GTCGTCAAT ATCGGGTGC GCGCGGCGAC GATGCGGTAA
40  551  TGTATTTCTT CGCACCGTCC CTGAACAATA TTCCGGCACA AATCGGCTAT
601  ACCGACGACG GCAAACCTA TACGCTGAAA CTCAAATCGG TGCAGATCAA
651  CCGCCAGGCA GCCAAACCGT AA

```

This corresponds to the amino acid sequence <SEQ ID 776; ORF120-1>:

```

1  MMKTFKNIFS AAILSAALPC AYAAGLPQSA VLHYSGSYGI PATMTFERSG
45  51  NAYKIVSTIK VPLYNIRFES GGTVVGNTLH PTYYRDIRRG KLYAEAKFAD
101  GSVTYGKAGE SKTEQSPKAM DLFTLAWQLA ANDAKLPPGL KITNGKKLYS
151  VGGLNKAGTG KYSIGGVETE VVKYRVRRGD DAVMYFFAPS LNNIPAQIGY
201  TDDGKTYTLK LKSVQINGQA AKP*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF120 shows 92.4% identity over a 184aa overlap with an ORF (ORF120a) from strain A of *N.*

*meningitidis*:

```

5          10          20          30
   orf120.pep          IPATMTFERSGNAYKIVSTIKVPLYNIRFE
   orf120a      SAAILSAAALPCAYAAGLPXSAVLHYSGSYGIPATXXXXXXXXNAXKIVSTIKVPLYNIRFE
10          10          20          30          40          50          60
   orf120.pep          40          50          60          70          80          90
   orf120a      SGGTVVGNTLHPTYYRDIRRGKLYAEAKFADGSVITYGKAGESKTEQSPKAMD LFTLAWQL
15          70          80          90          100          110          120
   orf120.pep          SGGTVVGNTLHPTYYRDIRRGKLYAEAKFADGSVITYGKAXXXXXXQSPKAMD LFTLAWQL
   orf120a      SGGTVVGNTLHPTYYRDIRRGKLYAEAKFADGSVITYGKAXXXXXXQSPKAMD LFTLAWQL
20          100          110          120          130          140          150
   orf120.pep          AANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRRGDDAVMYFFAP
   orf120a      AANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRRGDDAVMYFFAP
25          130          140          150          160          170          180
   orf120.pep          SLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAPX
   orf120a      SLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAPX
30          160          170          180
   orf120.pep          SLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAPX
   orf120a      SLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAPX
35          190          200          210          220

```

The complete length ORF120a nucleotide sequence <SEQ ID 777> is:

```

30      1  ATGATGAAGA CTTTAAAAA TATATTTTCC GCGCCATT TGTCCGCCGC
      51  CCTGCCGTGC GCGTATGCGG CAGGGCTGCC CNAATCCGCC GTGCTGCACT
      101 ATTCCGGCAG CTACGGCATT CCGCCACNA NNANNTNNGN ACNNNGNGNC
      151 AATGCTTNC AATCGTTTC GACGATTA A GTGCCGCTAT ACAATATCCG
      201 TTTGAGTCC GCGGTACGG TTGTCGGCA TACCCTGCAC CCTACCTACT
      251 ATAGAGACAT ACGCAGGGGC AAATGTATG CGGAAGCCAA ATTCGCCGAC
35      301 GGCAGCGTAA CCTACGGCAA AGCGNNNNN ANCNNNNNNG NGCAAAGCCC
      351 CAAGGCTATG GATTTGTTC CGCTTGCNTG GCAGTTGGCG GCAATGACG
      401 CGAAACTCCC CCGGGGCTG AAAATCACC ACGGCAAAA ACTTTATTCC
      451 GTCGGCGGTT TGAATAAGGC GGTACAGGA AAATACAGCA TAGCGCGCGT
      501 GGAAACCGAA GTCGTCAAT ATCGGGTGCG GCGCGGCGAC GATCGCGTAA
40      551 TGTATTCTT CGCACCGTCC CTGAACAATA TTCCGGCACA AATCGGCTAT
      601 ACCGACGACG GCAAACCTA TACGCTGAAA CTCAAATCGG TGCAGATCAA
      651 CGGCCAGGCA GCCAAACCGT AA

```

This encodes a protein having amino acid sequence <SEQ ID 778>:

```

45      1  MMKTFKNIFS AAILSAAALPC AYAAGLPXSA VLHYSGSYGI PATXXXXXXX
      51  NAXKIVSTIK VPLYNIRFES GGTVVGNTLH PTYYRDIRRG KLYAEAKFAD
      101 GSVITYGKAX XXXXQSPKAM DLFTLAWQLA ANDAKLPPGL KITNGKKLYS
      151 VGGLNKAGTG KYSIGGVETE VVKYRVRRGD DAVMYFFAPS LNNIPAQIGY
      201 TDDGKTYTLK LKSVQINGQA AKP*

```

ORF120a and ORF120-1 show 93.3% identity in 223 aa overlap:

```

50          10          20          30          40          50          60
   orf120a.pep      MMKTFKNIFS AAILSAAALPCAYAAGLPXSAVLHYSGSYGIPATXXXXXXXXNAXKIVSTIK
   orf120-1          MMKTFKNIFS AAILSAAALPCAYAAGLPQSAVLHYSGSYGIPATMTFERSGNAYKIVSTIK
55          10          20          30          40          50          60
   orf120a.pep          70          80          90          100          110          120
   orf120a.pep      VPLYNIRFESGGTVVGNTLHPTYYRDIRRGKLYAEAKFADGSVITYGKAXXXXXXQSPKAM

```

5	orf120-1	VPLYNIRFESGGTVVGNTLHPTYYRDIRRGKLYAEAKFADGSVITYGKAGESKTEQSPKAM	70	80	90	100	110	120
	orf120a.pep	DLFTLAWQLAANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRRGD	130	140	150	160	170	180
10	orf120-1	DLFTLAWQLAANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRRGD	130	140	150	160	170	180
	orf120a.pep	DAVMYFFAPSLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAPX	190	200	210	220		
15	orf120-1	DAVMYFFAPSLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAPX	190	200	210	220		

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF120 shows 97.8% identity over 184 aa overlap with a predicted ORF (ORF120ng) from *N.gonorrhoeae*:

20	orf120.pep	I PATMTFERSGNAYKIVSTIKVPLYNIRFE	30
	orf120ng	S AAILSAALPCAYAARLPQSAVLHYSGSYGIPATMTFERSGNAYKIVSTIKVPLYNIRFE	69
25	orf120.pep	SGGT VVGNTLHPTYYRDIRRGKLYAEAKFADGSVITYGKAGESKTEQSPKAMD LFTLAWQL	90
	orf120ng	SGGT VVGNTLHPAYYKDIRRGKLYAEAKFADGSVITYGKAGESKTEQSPKAMD LFTLAWQL	129
30	orf120.pep	AANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRRGDDAVMYFFAP	150
	orf120ng	AANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRRGDDTVTYFFAP	189
35	orf120.pep	SLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAP	184
	orf120ng	SLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAP	223

The complete length ORF120ng nucleotide sequence <SEQ ID 779> is:

35	1	ATGATGAAGA	CTTTTAAAA	TATATTTTCC	GCCGCCATT	TGTCCGCCGC
	51	CCTGCCGTGC	GCGTATGCGG	CAAGGCTACC	CCAATCCGCC	GTGCTGCACT
40	101	ATTCCGCGCAG	CTACGGCATT	CCCGCCACGA	TGACATTGA	ACGCAGCGGC
	151	AATGCTTACA	AAATCGTTTC	GACGATTAAA	GTGCCGCTAT	ACAATATCCG
45	201	TTTCGAATCC	GGCGGTACGG	TTGTCGGCAA	TACCCTGCAC	CCTGCCTACT
	251	ATAAAGACAT	ACGCAGGGGC	AACTGTATG	CGGAAGCCAA	ATTCGCCGAC
50	301	GGCAGCGTAA	CCTACGGCAA	AGCGGGCGAG	AGCAAAACCG	AGCAAAGCCC
	351	CAAGGCTATG	GATTGTGTTCA	CGCTTGCTG	GCAAGTGCGC	GCAAATGACG
55	401	CGAAACTCCC	CCCGGGTCTG	AAAATCACC	ACGGCAAAA	ACTTTATTCC
	451	GTCGGCGGCC	TGAATAAGGC	GGGTACGGGA	AAATACAGCA	TaggCGGCGT
60	501	GGAAACCGAA	GTCGTCAAAT	ATCGGGTGCG	GCGCGGCGAC	GATACGGTAA
	551	CGTATTTCTT	CGCACCGTCC	CTGAACAATA	TTCCGGCACA	AATCGGCTAT
65	601	ACCGACGACG	GCAAAACCTA	TACGCTGAAG	CTCAAATCGG	TGCAGATCAA
	651	CGGACAGGCC	GCCAAACCGT	AA		

This encodes a protein having amino acid sequence <SEQ ID 780>:

50	1	MMKTFKNIFS	AAILSAALPC	AYAARLPQSA	VLHYSGSYGI	PATMTFERSG
	51	NAYKIVSTIK	VPLYNIRFES	GGTVVGNTLH	PAYYKDIRRG	KLYAEAKFAD
55	101	GSVITYGKAGE	SKTEQSPKAM	DLFTLAWQLA	ANDAKLPPGL	KITNGKKLYS
	151	VGGLNKAGTG	KYSIGGVETE	VVKYRVRRGD	DTVTYFFAPS	LNNIPAQIGY
60	201	TDDGKTYTLK	LKSVQINGQA	AKP*		

In comparison with ORF120-1, ORF120ng shows 97.8% identity in 223 aa overlap:

60	orf120-1.pep	MMKTFKNIFS AAILSAALPCAYAAGLPQSAVLHYSGSYGIPATMTFERSGNAYKIVSTIK	10	20	30	40	50	60
	orf120ng	MMKTFKNIFS AAILSAALPCAYAARLPQSAVLHYSGSYGIPATMTFERSGNAYKIVSTIK	10	20	30	40	50	60

		70	80	90	100	110	120
	orf120-1.pep	VPLYNIRFESGGTVVGN	TLHPTYYRDIRRGKLYAEAKFADG	SVTYGKAGESKTEQSPKAM			
5	orf120ng	VPLYNIRFESGGTVVGN	TLHPAYYKDIRRGKLYAEAKFADG	SVTYGKAGESKTEQSPKAM			
		70	80	90	100	110	120
		130	140	150	160	170	180
10	orf120-1.pep	DLFTLAWQLAANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE	VVKYRVRRGD				
	orf120ng	DLFTLAWQLAANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE	VVKYRVRRGD				
		130	140	150	160	170	180
		190	200	210	220		
15	orf120-1.pep	DAVMYFFAPSLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAPX					
	orf120ng	DTVTYFFAPSLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAPX					
		190	200	210	220		

This analysis, including the presence of a putative leader sequence in the gonococcal protein suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 93

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 781>:

1	ATGTATCGGA	GGAAAGGGCG	GGGCATCAAG	CCGTGGATGG	GTGCCGGTGC
51	.GCGTTTGCC	GCCTTGGTCT	GGCTGGTTTT	CGCGCTCGGC	GATACTTTGA
101	CTCCGTTTGC	GGTTGCGGCG	GTGCTGGCGT	ATGTATTGGA	CCCTTTGGTC
151	GAATGGTTGC	AGAAAAAGGG	TTTGAACCGT	GCATCCGCTT	CGATGCTGT
201	GATGGTGTTC	TCCTTGATTT	TGTTGTTGGC	ATTATTGTG	ATTATCGTCC
251	CTATGCTGGT	CGGGCAGTTC	AACAATTG	CATCGCGCCT	GCCCCAATTA
301	ATCGGTTTTC	TGCAGAACAC	GCTGCTGCCG	TGTTGAAAA	ATACAATCGG
351	CGGATATGTG	GAAATCGATC	AGGCATCTAT	TATTGCGTGG	CTTCAGGCGC
401	ATACGGGAGA	GTTGAGCAAC	GCGCTTAAGG	CGTGGTTTCC	CGTTTTGATG
451	AGGCAGGGCG	GCAATATT..			

This corresponds to the amino acid sequence <SEQ ID 782; ORF121>:

1	MYRRKGRGIK	PWMGAGXAFA	ALVWLVFALG	DTLTPFAVAA	VLAYVLDPLV
51	EWLQKKGLNR	ASASMSVMVF	SLILLALLL	IIVPMLVGQF	NNLASRLPQL
101	IGFMQNTLLP	WLKNTIGGYV	EIDQASIIAW	LQHTGELSN	ALKAWFPVLM
151	RQGGNI..				

Further work revealed the complete nucleotide sequence <SEQ ID 783>:

1	ATGTATCGGA	GGAAAGGGCG	GGGCATCAAG	CCGTGGATGG	GTGCCGGTGC
51	GGCGTTTGCC	GCCTTGGTCT	GGCTGGTTTT	CGCGCTCGGC	GATACTTTGA
101	CTCCGTTTGC	GGTTGCGGCG	GTGCTGGCGT	ATGTATTGGA	CCCTTTGGTC
151	GAATGGTTGC	AGAAAAAGGG	TTTGAACCGT	GCATCCGCTT	CGATGCTGT
201	GATGGTGTTC	TCCTTGATTT	TGTTGTTGGC	ATTATTGTG	ATTATCGTCC
251	CTATGCTGGT	CGGGCAGTTC	AACAATTG	CATCGCGCCT	GCCCCAATTA
301	ATCGGTTTTC	TGCAGAACAC	GCTGCTGCCG	TGTTGAAAA	ATACAATCGG
351	CGGATATGTG	GAAATCGATC	AGGCATCTAT	TATTGCGTGG	CTTCAGGCGC
401	ATACGGGAGA	GTTGAGCAAC	GCGCTTAAGG	CGTGGTTTCC	CGTTTTGATG
451	AGGCAGGGCG	GCAATATTGT	CAGCAGTATC	GGCAACCTGC	TGCTGCTTCC
501	CTTGCTGCTT	TACTATTTCC	TGCTGGATTG	GCAGCGGTGG	TCGTGCGGCA
551	TTGCCAAACT	GGTCCGAGG	CGTTTTGCCG	GTGCTTATAC	GCGCATTACA
601	GGCAATTGGA	ACGAGGTATT	GGGCGAATT	TTGCGCGGGC	AGCTTCTGGT
651	AATGCTGATT	ATGGGCTTGG	TTTACGGTTT	GGGATTGGTG	CTGGTCGGGC
701	TGGATTCCGG	GTTTGCCATC	GGTATGCTTG	CCGGTATTTT	GGTGTGTTGC
751	CCTTATCTCG	GGCGGTTTAC	GGGATTGCTG	CTTGCCACCG	TCGCCGCCTT
801	GCTCCAGTTC	GGTTCGTGGA	ACGGCATCCT	ATCGGTTTGG	GCGGTTTTTG
851	CCGTAGGACA	GTTTCTCGAA	AGTTTTTTCA	TTACGCCGAA	AATCGTGGGA

901 GACCGTATCG GGCTGTCGCC GTTTTGGGTT ATCTTTTCGC TGATGGCGTT  
 951 CGGGCAGCTG ATGGGCTTTG TCGGAATGTT GCGGGGATTG CCTTTGGCCG  
 1001 CCGTAACCTT GGTCTTGCTT CCGAGGGCG TGCAGAAATA TTTTCCCGCG  
 1051 AGTTTTTACC GGGGCAGGTA G

5 This corresponds to the amino acid sequence <SEQ ID 784; ORF121-1>:

1 MYRRKGRGIK PWMGAGAAFA ALVWLVFALG DTLTPFAVAA VLAYVLDPLV  
 51 EWLQKKGLNR ASASMSVMVF SLILLALLL IIVPMLVGQF NNLASRLPQL  
 101 IGFQMNTLLP WLKNTIGGYV EIDQASIIAW LQAHTGELSN ALKAWFPVLM  
 151 RQGGNIVSSI GNLLLLPLLL YYFLLDWQRW SCGIAKLVPR RFAGAYTRIT  
 10 201 GNLNEVLGEF LRGQLLVMLI MGLVYGLGLV LVGLDSGFAI GMLAGILVEV  
 251 PYLGAFGLL LATVAALLQF GSWNGILSVW AVFAVGQFLE SFFITPKIVG  
 301 DRIGLSPFWV IFSLMAFGQL MGFVGMLAGL PLAAVTLVLL REGVQKYFAG  
 351 SFYRGR\*

Computer analysis of this amino acid sequence gave the following results:

15 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF121 shows 98.7% identity over a 156aa overlap with an ORF (ORF121a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
20	orf121.pep	MYRRKGRGIK	PWMGAGXAF	ALVWLVFALG	DTLTPFAVA	AVLAYVLDPL	VEWLQKKGLNR
	orf121a	MYRRKGRGIK	PWMDAGAAFA	ALVWLVFALG	DTLTPFAVA	AVLAYVLDPL	VEWLQKKGLNR
		10	20	30	40	50	60
25	orf121.pep	ASASMSVMVF	SLILLALLL	IIVPMLVGQF	NNLASRLPQL	LIGFMQNTLL	PWLKNTIGGYV
	orf121a	ASASMSVMVF	SLILLALLL	IIVPMLVGQF	NNLASRLPQL	LIGFMQNTLL	PWLKNTIGGYV
		70	80	90	100	110	120
30	orf121.pep	EIDQASIIAW	LQAHTGELSN	ALKAWFPVLM	RQGGNI		
	orf121a	EIDQASIIAW	LQAHTGELSN	ALKAWFPVLM	RQGGNIVSS	IGNLLLLPL	LLYYFLLDWQRW
		130	140	150	160	170	180
35	orf121a	SCGIAKLVPR	RFAGAYTRIT	GNLNEVLGEF	LRGQLLVML	IMGLVYGLGL	VLVGLDSGFAI
		190	200	210	220	230	240

The complete length ORF121a nucleotide sequence <SEQ ID 785> is:

40 1 ATGTATCGGA GGAAAGGCG GGCATCAAG CCGTGGATGG ATGCCGGTGC  
 51 GGCGTTTGCC GCCTTGGTCT GGCTGGTTT CCGCTCGGC GATACTTTGA  
 101 CTCCGTTTGC GGTGCGGCG GTGCTGGCGT ATGTATTGGA CCTTTGGTC  
 151 GAATGGTTGC AGAAAAAGG TTTGAACCGT GCATCCGCTT CGATGTCTGT  
 201 GATGGTGTTT TCCTTGATTT TGTGTTGGC ATTATTGTTG ATTATTGTCC  
 251 CTATGCTGGT CGGGCAGTTC AACAAATTGG CATCGGCCTT GCCCAATTA  
 301 ATCGGTTTGA TGCAGAACAC GCTGCTGCCG TGGTTGAAAA ATACAATCGG  
 351 CGGATATGTG GAAATCGATC AGGCATCTAT TATTGCGTGG CTTCAGGCGC  
 401 ATACGGGCGA GTTGAGCAAC GCGCTTAAGG CGTGGTTTCC CGTTTGATG  
 451 AGGCAGGGCG GCAATATTGT CAGCAGTATC GGCAACCTGC TCGTGCTTCC  
 501 CTTGCTGCTT TACTATTTC TGTGCTGATTG GCAGCGGTGG TCGTGCGGCA  
 551 TTGCCAACT GGTCCGAGG CGTTTTCGCG GTGCTTATAC GCGCATTACA  
 601 GGCAATTTGA ACGAGGTATT GGGCGAATTT TTGCGCGGGC AGCTTCTGGT  
 651 GATGCTGATT ATGGGTTTGG TTACGGCTT GGGGTGTTG CTGGTCGGGC  
 701 TGGATTCGGG GTTTCGAATC GGTATGGTTG CCGGTATTTT GGTTTTGT  
 751 CCTATTGG GCGCGTTTAC AGGACTGCTG CTGGCAACCG TCGCCGCCTT  
 801 GCTCCAGTTC GGTTCGTGGA ACGGCATCTT GGCTGTTTGG GCGGTTTTTG  
 851 CCGTAGGACA GTTCTCGAA AGTTTTTCA TTACGCCGAA AATCGTGGGA  
 901 GACCGTATCG GCCTGTCGCC GTTTGGGTT ATCTTTTCGC TGATGGCGTT  
 951 CGGGCAGCTG ATGGGCTTTG TCGGAATGTT GCGCGGATTG CCTTTGGCCG  
 1001 CCGTAACCTT GGTCTTGCTT CCGAGGGCG TGCAGAAATA TTTTCCCGCG  
 60 1051 AGTTTTTACC GGGGCAGGTA G

This encodes a protein having amino acid sequence <SEQ ID 786>:

```

1  MYRRKGRGIK PWMDAGAAFA ALVWLVFALG DTLTPFAVA VLAYVLDPLV
51 EWLQKKGLNR ASASMSVMVF SLILLALLL IIVPMLVGQF NNLASRLPQL
101 IGFMQNTLLP WLKNTIGGYV EIDQASIIAW LQAHTGELSN ALKAWFPVLM
151 RQGGNIVSSI GNLLLLPLLL YYFLLDWQRW SCGIAKLVPR RFAGAYTRIT
201 GNLNEVLGEF LRGQLLVMLI MGLVYGLGLV LVGLDSGFAI GMVAGILVFE
251 PYLGAFTGLL LATVAALLQF GSWNGILAVW AVFAVGQFLE SFFITPKIVG
301 DRIGLSPFWV IFSLMAFGQL MGFVGMLAGL PLAATLVLL REGVQKYFAG
351 SFYRGR*

```

10 ORF121a and ORF121-1 show 99.2% identity in 356 aa overlap:

```

15 orf121a.pep      10      20      30      40      50      60
    MYRRKGRGIKPWMDAGAAFAALVWLVFALGDTLTPFAVAAYVLDPLVEWLQKKGLNR
    |||||
orf121-1           10      20      30      40      50      60
    MYRRKGRGIKPWMDAGAAFAALVWLVFALGDTLTPFAVAAYVLDPLVEWLQKKGLNR
    |||||

20 orf121a.pep      70      80      90      100     110     120
    ASASMSVMVFSLILLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV
    |||||
orf121-1           70      80      90      100     110     120
    ASASMSVMVFSLILLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV
    |||||

25 orf121a.pep      130     140     150     160     170     180
    EIDQASIIAWLQAHTGELSNALKAWFPVLMRQGGNIVSSIGNLLLLPLLLYYFLLDWQRW
    |||||
orf121-1           130     140     150     160     170     180
    EIDQASIIAWLQAHTGELSNALKAWFPVLMRQGGNIVSSIGNLLLLPLLLYYFLLDWQRW
    |||||

30 orf121a.pep      190     200     210     220     230     240
    SCGIAKLVPRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLVGLVGLDSGFAI
    |||||
orf121-1           190     200     210     220     230     240
    SCGIAKLVPRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLVGLVGLDSGFAI
    |||||

35 orf121a.pep      250     260     270     280     290     300
    GMVAGILVFPYLYLGAFTGLLATVAALLQFGSWNGILAVWAVFAVGQFLESFFITPKIVG
    ||:|||||
orf121-1           250     260     270     280     290     300
    GMLAGILVFPYLYLGAFTGLLATVAALLQFGSWNGILSVWAVFAVGQFLESFFITPKIVG
    ||:|||||

40 orf121a.pep      310     320     330     340     350
    DRIGLSPFWVIFSMAFGQLMGFVGMLAGLPLAATLVLLREGVQKYFAGSFYRGRX
    |||||
orf121-1           310     320     330     340     350
    DRIGLSPFWVIFSMAFGQLMGFVGMLAGLPLAATLVLLREGVQKYFAGSFYRGRX
    |||||

45 orf121a.pep      310     320     330     340     350
    DRIGLSPFWVIFSMAFGQLMGFVGMLAGLPLAATLVLLREGVQKYFAGSFYRGRX
    |||||
orf121-1           310     320     330     340     350
    DRIGLSPFWVIFSMAFGQLMGFVGMLAGLPLAATLVLLREGVQKYFAGSFYRGRX
    |||||

```

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF121 shows 97.4% identity over a 156 aa overlap with a predicted ORF (ORF121ng) from *N.gonorrhoeae*:

```

50 orf121.pep      MYRRKGRGIKPWMDAGAAFAALVWLVFALGDTLTPFAVAAYVLDPLVEWLQKKGLNR      60
    |||||
orf121ng          MYRRKGRGIKPWMDAGAAFAALVWLVFALGDTLTPFAVAAYVLDPLVEWLQKKGLNR      60
    |||||

55 orf121.pep      ASASMSVMVFSLILLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV      120
    |||||
orf121ng          ASASMSVMVFSLILLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV      120
    |||||

orf121.pep      EIDQASIIAWLQAHTGELSNALKAWFPVLMRQGGNI      156
    |||||
orf121ng          EIDQASIIAWLQAHTGELSNALKAWFPVLMKQGGNIVSTIGNLLLLPLLLYYFLLDWHRW      180
    |||||

```

An ORF121ng nucleotide sequence <SEQ ID 787> was predicted to encode a protein having amino acid sequence <SEQ ID 788>:

```

1 MYRRKGRGIK PWMGAGAAFA ALVWLIVYALG DTLTPFAVAA VLAYVLDPLV
51 EWLQKKGLNR ASASMSVMVF SLILLALLL IIVPMLVGQF NNLASRLPQL
101 IGFMQNTLLP WLKNTIGGYV EIDQASIIAW FQHTGELSN ALKAWFPVLM
151 KQGGNIVSTI GNLLLPPLL YYFLLDWHRW SCGIPKLVPR RFAGAYTRIT
201 GNLNVWVKF LRGQLGETE RGAIVCRVGR ECWEGGARS RPSDDGWPRW
251 GGG*

```

Further work revealed the following gonococcal DNA sequence <SEQ ID 789>:

```

10 1 ATGTATCGGA GAAAAGGACG GGGCATCAAG CCGTGGATGG GTGCCGGCGC
51 GGCCTTTGCC GCCTTGGTCT GGCTGGTTA CGCGCTCGGC GATACTTTGA
101 CTCCGTTTGC GGTTCGCGCG GTGCTGGCGT ATGTGTTGGA CCCTTTGGTC
151 GAATGGTTGC AGAAAAGGG TTTGAACCGT GCATCCGCTT CGATGCTCTG
15 201 GATGGTGTTC TCCTTGATTT TGTTGTTGGC ATTATTGTTG ATTATTGTCC
251 CTATGCTGGT CCGGCAGTTC AATAATTGG CATCTCGCTT GCCCAATTA
301 ATCGGTTTTC TGCAGAACAC GCTGCTGCCG TGGTTGAAAA ATACAATCGG
351 CGGATATGTG GAAATCGATC AGGCATCTAT TATTGCGTGG TTTCAGGCGC
401 ATACGGGCGA GTTGAGCAAC GCGCTTAAGG CGTGGTTTCC CGTTTGTATG
451 AAACAGGGCG GCAATATTGT CAGCAGTATC GGCAACCTGC TGCTGCCGCC
20 501 CTTGCTGCTT TACTATTTC TGCTGGATTG GCAGCGGTGG TCGTGCGGCA
551 TCGCCAACT GGTTCGAGG CGTTTTCGCG GTGCTTATAC GCGCATTACG
601 GGTAATTTGA ACGAGGTATT GGGCGAATTT TTGCGCGGTC AGCTTCTGGT
651 GATGCTGATT ATGGGCTTGG TTTACGGTTT GGGATTGATG CTAGTCGGAC
701 TGGATTCCGG ATTTGCCATC GGTATGTTG CCGGTATTTT GGTGTTTGT
25 751 CCCTATTTGG GTGCGTTTAC GGGATTGCTG CTTGCCACTG TTGCAGCCTT
801 GCTCCAGTTC GGTTCGTGGA ACGGAATCTT GGCTGTTTGG GCGGTTTTTG
851 CCGTCGGTCA GTTCTCGAA AGTTTTTTCA TTACGCCGAA AATTGTAGGA
901 GACCGTATCG GCCTGTGCGC GTTTTGGGTT ATCTTTTCGC TGATGGCGTT
951 CGGAGAGCTG ATGGGCTTTG TCGGAATGTT GGCCGGATTG CCTTTGGCCG
30 1001 CCGTAACCTT GGTCTTGCTT CCGGAGGGCG CGCAGAAATA TTTTGCCGGC
1051 AGTTTTTACC GGGGCAGGTA G

```

This corresponds to the amino acid sequence <SEQ ID 790; ORF121ng-1>:

```

1 MYRRKGRGIK PWMGAGAAFA ALVWLIVYALG DTLTPFAVAA VLAYVLDPLV
35 51 EWLQKKGLNR ASASMSVMVF SLILLALLL IIVPMLVGQF NNLASRLPQL
101 IGFMQNTLLP WLKNTIGGYV EIDQASIIAW FQHTGELSN ALKAWFPVLM
151 KQGGNIVSSI GNLLLPPLL YYFLLDWQRW SCGIKLVPR RFAGAYTRIT
201 GNLNEVLGEF LRGQLVMLI MGLVYGLGLM LVGLDSGF AI GMVAGILVEF
251 PYLGAFTGLL LATVAALLQF GSWNGILAVW AVFAVGQFLE SFFITPKIVG
40 301 DRIGLSPFWV IFSLMAFGEI MGFVGMLAGL PLAAVTLVLL REGAQKYFAG
351 SFYRGR*

```

ORF121ng-1 and ORF121-1 show 97.5% identity in 356 aa overlap:

```

10 20 30 40 50 60
orfl21-1.pep MYRRKGRGIKPWMGAGAAFAALVWLIVFALGDTLTPFAVAAVLAYVLDPLVEWLQKKGLNR
45 orfl21ng-1 MYRRKGRGIKPWMGAGAAFAALVWLIVYALGDTLTPFAVAAVLAYVLDPLVEWLQKKGLNR
10 20 30 40 50 60
70 80 90 100 110 120
orfl21-1.pep ASASMSVMVFSLILLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV
50 orfl21ng-1 ASASMSVMVFSLILLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV
70 80 90 100 110 120
130 140 150 160 170 180
orfl21-1.pep EIDQASIIAWLQHTGELSNALKAWFPVLMRQGGNIVSSIGNLLLPPLLYYFLLDWQRW
55 orfl21ng-1 EIDQASIIAWFQHTGELSNALKAWFPVLMKQGGNIVSSIGNLLLPPLLYYFLLDWQRW
130 140 150 160 170 180
60 190 200 210 220 230 240
orfl21-1.pep SCGIKLVPRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLVGLDSGF AI

```



```

      |||
orfl21ng-1 SCGI AKLV PRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLMLVGLDSGFAI
           190      200      210      220      230      240
5
      250      260      270      280      290      300
orfl21-1.pep GMLAGILVFPYPLGAFTGLLLATVAALLQFGSWNGILSVWAVFAVGQFLESFFITPKIVG
      |||
orfl21ng-1 GMVAGILVFPYPLGAFTGLLLATVAALLQFGSWNGILAVWAVFAVGQFLESFFITPKIVG
           250      260      270      280      290      300
10
      310      320      330      340      350
orfl21-1.pep DRIGLSPFWVIFSLMAFGQLMGFVGMLAGLPLAAVTLVLLREGVQKYFAGSFYRGRX
      |||
orfl21ng-1 DRIGLSPFWVIFSLMAFGELMGFVGMLAGLPLAAVTLVLLREGAQKYFAGSFYRGRX
           310      320      330      340      350
15

```

In addition, ORF121ng-1 shows homology to a permease from *H.influenzae*:

```

sp|P43969|PERM HAEIN PUTATIVE PERMEASE PERM HOMOLOG Length = 349
Score = 69.9 bits (168), Expect = 2e-11
Identities = 67/317 (21%), Positives = 120/317 (37%), Gaps = 7/317 (2%)
20
Query: 26 VYALGDTLTPFAVAAYVLDPVLEWL-QKKGLNRASASMSVMVFSXXXXXXXXXXVP 84
      +Y GD + P +A VL+Y+L+ + +L Q R A++ + VP
Sbjct: 32 IYFFGDLIAPLLIALVLSYLLEIPINFLNQYLKCPRLATILIFGSFIGLAAVFFLVLP 91
25
Query: 85 MLVGQFNNLASRLPOLIGFMQNTLLPWLKNTIGGYE-IDQASIIAWFQAHTGELSNAK 143
      ML Q +L S LP + N WL N Y E ID + + + F + ++ +
Sbjct: 92 MLWNQITISLLSDLPAMF----NKSNEWLLNLPKNYPELIDYSMVDSIFNSVREKILGFGE 147
Query: 144 AWFVLMKQGGNIVSSIGNXXXXXXXXXXXXXWQWSCGI AKLV PRRFAGAYTRITGNL 203
30
      + + + N+VS D G+++ +P+ A+ R +
Sbjct: 148 SAVKLSLASIMNLVSLGIYAFLVPLMMFFMLKDKSELLQGVSRFLPKNRNLAFXRWK-EM 206
Query: 204 NEVLGEFLRGQXXXXXXXXXXXXXXXXXXXXXDSGFAIGMVAGILVFPYXXXXXXXXXXXX 263
      + + ++ G+ + + G+ V VPY
35
Sbjct: 207 QQQISNYIHGKLEILIVTLITYIIFLIFGLNYPDLLAFVGLSVLVPYIGAVIVTIPVA 266
Query: 264 XXXXXQFGSWNGILAVWAVFAVGQFLESFFITPKIVGDRIGLSPFWVIFSLMAFGELMGF 323
      QFG + FAV Q L+ + P + + + L P +I S++ FG L GF
40
Sbjct: 267 LVALFQFGISPTFWYIIIAFAVSQLLDGNLLVPYLFSEAVNLHPLIIIIISVLIFGGLWGF 326
Query: 324 VGMLAGLPLAAVTLVLL 340
      G+ +PLA + ++
Sbjct: 327 WGVFFAIPLATLVKAVI 343

```

Based on this analysis, including the presence of a putative leader sequence and transmembrane domains in the two proteins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 94

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 791>:

```

50      1 ..ACTGCTTTT CGGCGGCGCT GCGCTTGAGT CCATCATGAC TCGTCATATT
      51 TTTGTCTCTT GGGAAACCGT ATCAACAAAC AGCCGCCATC TTAACATTTT
      101 TTTGCACGTC CTGCCGCGCG CGTTCAAATG CGTACCAGCA ATACCGCCGC
      151 CTGCGCCTCT ATGCCTTCCA TCCGCCCGAG ATAGCCGAGT TTTTCGTGG
55      201 TTTTGCTTTT GATGTTGACG CACGAAATGT CTATGCCCAA ATCGGCGGCG
      251 ATGTTGGCAC GCATTGCGG AATGTGCGGC GCGAGTGTGG GTTCTGTGTC
      301 AATCACGTC GTATCGACAT TGACCGCTG CCAACCCTGC GCCTGAACGC
      351 TTTGATACGC CGCACGCAA AGGACGCGGC TGTCCGCATC TTTGAACTCT
      401 GCGGCGGTGT CGGGGAAATG GCTGCCGATA TCGCCCAAAC CTGCCGCACC

```

451 GAGCAGCGCG TCGGTAACGG CGTGCAGCAG CGCATCGGCA TCGGAGTGTC  
501 CGAGCAGCCC TTTTCAAAT GGGATTCAA CTCCGCCAAG TATCAG..

This corresponds to the amino acid sequence <SEQ ID 792; ORF122>:

5 1 ..TAFSAALRLS PSXLVIFLSF GKPYQQTAAI LTFECTSCPP RSNAYQQYRR  
51 LRLYAFHPPE IAEFFVGFAF DVDARNVYAO IGGDVGTHLR NVRRECGFLC  
101 NHGRIDIDRL PTLRLNALIR RTQKDAAVRI FELCGGVGEM AADIAQTCRT  
151 EQRVGNVQOQ RIGIGVSEOP FFKWDFNSAK YQ..

Further work revealed the complete nucleotide sequence <SEQ ID 793>:

10 1 ATATCGTACT GGGCAAGCAG TTCGCCGGAT TTTTGGGAAG TAGATACCGC  
51 GCCTTTGATT TTTTGGCCGC TCTTACCCAA GGCTTCGATG AAAAAAGTTGA  
101 TGGTCGAGCC GGTACCGATG CCGATATATT CATTTCGGG TACGAATTCG  
151 ACTGCTTTTT CGGCGGCGAT GCGCTTGAGT TCGTCTTG TGCTCATATT  
201 TTTGTCTTTT GGGAAACCGT ATCAACAAAC AGCCGCCATC TTAACATTTT  
15 251 TTTGCACGTC CTGCCCCGCG CGTTCAAATG CGTACCAGCA ATACCGCCGC  
301 CTGCGCCTCT ATGCCTTCCA TCCGCCGAG ATAGCCGAGT TTTTCGTTGG  
351 TTTTGCCTTT GATGTTGACG CACGAAATGT CTATGCCCAA ATCGGCGGCG  
401 ATGTTGGCAC GCATTTCGGG AATGTGCGGC GCGAGTTTGG GTTTCTGTGC  
451 AATCAGGGTC GTATCGACAT TGACCGCCTG CCAACCCTGC GCCTGAACGC  
501 TTTGATACGC CGCACGCAA AGGACGCGGC TGTCCGCATC TTTGAACCTC  
20 551 CGGCGGCTGT CGGGGAAATG GCTGCCGATA TCGCCCAAAC CTGCCGCACC  
601 GAGCAGCGCG TCGGTAACGG CGTGCAGCAG CGCATCGGCA TCGGAGTGTC  
651 CGAGCAGCCC TTTTCAAAT GGGATTCAA CTCCGCCAAG TATCAGCTTT  
701 CTGCCTTCG TCAGTTGGTG GACATCGTAG CCCTGTCCGA TACGGATGTT  
751 CGTCATCGTT TGTGTTCTG A

25 This corresponds to the amino acid sequence <SEQ ID 794; ORF122-1>:

1 ISYWASSSPD FLEVDTAPLI FLPLLKASM KLMVEPVPM PIYSFSGTNS  
51 TAFSAAMRLS SSCVVIPLSF GKPYQQTAAI LTFECTSCPP RSNAYQQYRR  
101 LRLYAFHPPE IAEFFVGFAF DVDARNVYAO IGGDVGTHLR NVRREFGFLC  
151 NHGRIDIDRL PTLRLNALIR RTQKDAAVRI FELCGGVGEM AADIAQTCRT  
30 201 EQRVGNVQOQ RIGIGVSEOP FFKWDFNSAK YQLSAFGQLV DIVALSDDTV  
251 RHRLCS\*

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF122 shows 94.0% identity over a 182aa overlap with an ORF (ORF122a) from strain A of *N. meningitidis*:

35

orf122.pep				10	20	30
				TAFSAALRLS	PSXLVIFLSF	GKPYQQTAAI
				:	:	
orf122a	FLPLLKASM	KLMVEPVMP	MYSFSGTN	STAFSAAMRL	SSSCVVIPL	SFGKPYQQTAAI
40		30	40	50	60	70
orf122.pep		40	50	60	70	80
		LTFECTSCPP	RSNAYQQYRR	LRLYAFHPPE	IAEFFVGFAF	DVDARNVYAO
45	orf122a	LTFEXTSCPP	RSNPNYQQYRR	LRLYAFHAP	EITEFFVGFAF	XVDARNVYAO
		90	100	110	120	130
orf122.pep		100	110	120	130	140
		NVRRECGFLC	NHGRIDIDRL	PTLRLNALIR	RTQKDAAVRI	FELCGGVGEM
		:				
50	orf122a	NMRREFGFLC	NHGRIDIDRL	PTLRLNALIR	RTQKDAAVRI	FELCGGVGEM
		150	160	170	180	190
orf122.pep		160	170	180		
		EQRVGNVQOQ	RIGIGVSEOP	FFKWDFNSAK	YQ	
55	orf122a	EQRVGNVQOQ	RIGIGVSEOP	FFKWDFNSAK	YQLSAFGQLV	DIVALSDDTV
						RHRLCSX

210 220 230 240 250

The complete length ORF122a nucleotide sequence <SEQ ID 795> is:

```

1  ATATCATATT GGGCAAGCAG TTCACTGGAT TTTTGGGAAG TAGATACCGC
5  51  GCCTTTGATT TTTTGGCCG TCTTACCCAA GGCTTCGATG AAAAAGTTGA
101 TGGTCGAACC GGTACCGATG CCGATGTATT CGTTTTCGGG TACGAATTCG
151 ACTGCNTTTT CGGCGGCGAT GCGCTTGAGT TCGTCTTGTT TCGTCATATT
201 TTTGTCCTTT GGGAAACCGT ATCAACAAAC AGCCGCCATC TTAACATTTT
251 TNNNACGTC CTGCCCGCCG CGTTCAAATC CTTACCAGCA ATACCGCGC
301 CTGCGACTCT ATGCCTTCCA TGCGCCCGAG ATAACCGAGT TTTTCGTTGG
10  351 TTTTGCCTTT GANGTTGACG CACGAAATGT CTATGCCAA ATCGGCGGCG
401 ATGTTGGCAC GCATTTGCGG AATATGCGGC GCGAGTTTGG GTTCTGTGTC
451 AATCACGGTC GTATCGACAT TGACCGCCTG CCAACCTGCG GCCTGAACGC
501 TTTGATACGC CGCACGCAAA AGGACGCGGC TGTCCGCATC TTTGAACCTC
551 GCGGCGGTGT CGGGGAAATG GCTGCCGATA TCGCCCAAC CTGCCGACCC
15  601 GAGCAGCGCG TCGGTAACGG CGTGCAGCAG CGCATCGGCA TCGGAGTGTC
651 CGAGCAGCCC TTTTCAAAT GGGATTCAA CTCGCCAAG TATCAGCTTT
701 CTGCCTTCGG TCAGTTGGTG GACATCGTAG CCCTGTCCGA TACGGATGTT
751 CGTCATCGTT TGTGTTCTCG A

```

This encodes a protein having amino acid sequence <SEQ ID 796>:

```

20 1  ISYWASSSLD FLEVDTAPLI FLPLLPKASM KKLMEVPVPM PMYSFSGTNS
51  TAFSAAMRLS SSCVVIFLSF GKPYQQTAAI LTFEFTSCFP RSNPYQQYRR
101 LRLYAFHAP ETEFFVGFAF XVDARNVYAO IGGDVGTHLR NMRREFGFLC
151 NHGRIDIDRL PTLRLNALIR RTQKDAAVRI FELCGGVGEM AADIAQTCRT
201 EQRVGNVQO RIGIGVSEQP FFKWDFNSAK YQLSAFGQLV DIVALSDTDV
25  251 RHRLCS*

```

ORF122a and ORF122-1 show 96.9% identity in 256 aa overlap:

```

30 orf122a.pep 10 20 30 40 50 60
ISYWASSSLDFLEVDTAPLIFLPLLPKASMKKLMVEVPVPMPIYSFSGTNSTAFSAAMRLS
||||||| ||||||| ||||||| ||||||| ||||||| |||||||
orf122-1 ISYWASSSPDFLEVDTAPLIFLPLLPKASMKKLMVEVPVPMPIYSFSGTNSTAFSAAMRLS
10 20 30 40 50 60

35 orf122a.pep 70 80 90 100 110 120
SSCVVIFLSFGKPYQQTAAILTFFXTSCPPRSNPYQQYRRRLRLYAFHAPETEEFFVGFAF
||||||| ||||||| ||||||| ||||||| ||||||| |||||||
orf122-1 SSCVVIFLSFGKPYQQTAAILTFFXTSCPPRSNAYQQYRRRLRLYAFHPPEIAEFFVGFAF
70 80 90 100 110 120

40 orf122a.pep 130 140 150 160 170 180
XVDARNVYAOIGGDVGTHLRNMRREFGFLCNHGRIDIDRLPTLRLNALIRRTQKDAAVRI
||||||| ||||||| ||||||| ||||||| ||||||| |||||||
orf122-1 DVDARNVYAOIGGDVGTHLRNMRREFGFLCNHGRIDIDRLPTLRLNALIRRTQKDAAVRI
130 140 150 160 170 180

45 orf122a.pep 190 200 210 220 230 240
FELCGGVGEMAADIAQTCRTEQRVGNVQQRIGIGVSEQPFKWDFNSAKYQLSAFGQLV
||||||| ||||||| ||||||| ||||||| ||||||| |||||||
orf122-1 FELCGGVGEMAADIAQTCRTEQRVGNVQQRIGIGVSEQPFKWDFNSAKYQLSAFGQLV
190 200 210 220 230 240

50 orf122a.pep 250
DIVALSDTDVRHRLCSX
|||||||
orf122-1 DIVALSDTDVRHRLCSX
250
55

```

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF122 shows 89.6% identity over a 182 aa overlap with a predicted ORF (ORF122ng) from *N.gonorrhoeae*:

	orf122.pep	TAFSAALRLSPSXLVIFLSFGKPYQQTAAI	30
	orf122ng	FLPLLPKASMKKLMVEPVMPMYSFSGTNSTAFSAAMRLSSSCVVIFLSFGKPYQQTAAI	80
5	orf122.pep	LTFECTSCPPRSNAYQQYRRLRLYAFHPPEIAEFFVGFAFDVDARNVYAQIGGDVGTHLR	90
	orf122ng	LTFECTSWPPRSNPNYQQYRRLRLYAFHPPEIAEFFVGFAFDIDARNIDTQIGGDVGTHLR	140
10	orf122.pep	NVRRECGFLCNHGRIDIDRLPTLRLNALIRRTQKDAAVRIFELCGGVGEMAADIAQTCRT	150
	orf122ng	NVRCEFGFLCNHGRIDIDHLPRLNALIRRTQKDAAVRIFELCGGVGKMAADVAQTCRT	200
	orf122.pep	EQRVGNVQQRIGIGVSEQPFFKWFDFNSAKYQ	182
15	orf122ng	EQRVGNVQQRVGIRMPEQPFFKWFDFNSAKYQLSAFGQLVDIVALSDDIRHRLCS	256

The complete length ORF122ng nucleotide sequence <SEQ ID 797> is:

1	ATGTCGTACC	GGGCAAGCAG	TTCGCCGGAT	TTTTTGGAGG	TGAAACCGC
51	GCCTTTGATT	TTTTTACCGC	TTTGCCCAA	GGCTTCGATG	AAGAAATTGa
101	tgGTCGAACC	GgtacCGATG	CCGATGTATT	CGTTTTCGGG	TACGAATTCG
151	ACTGCTTTT	CGGCGGCGAT	GCGCttgAgt	TCgtcttgcg	TcgTCATATT
201	TTTAtccttt	gGAAaccct	atcaAcaAAc	agccgccatC	TTAACATTTT
251	TTTGACGtc	ctggccgcg	cgttcaAATc	cgtaccaGca	ataccgccgc
301	ctgcgcctCT	AtgcCTTCCA	TCCGCCCGAG	ATAGCCGAGT	TTTTCGTTGG
351	TTTGCCTTT	GATattGACG	CACGAAATAT	CGataCCAA	atcgccgCGC
401	ATGTGGCAC	GCATTTGCGG	AATGTGCGGT	GCGAGTTTGG	GTTTCTGTGC
451	AATCACGGTC	GTATCGACAT	TGACCACCTG	CCAACCCTGC	GCCTGAACGC
501	TTTGATACGC	CGCACGCAAA	AGGACGCGGC	TGTCCGCATC	TTTGAACCTCT
551	GCGGCGGTGT	CGGAAAAATG	GCTGCCGATG	TCGCCCAAAC	CTGCCGCACC
601	GAGCAGCgcy	tcggtaaCGG	CGTGACGAG	cgcgTcgGCA	TCCGAATGCC
651	CGAGCAGCCC	TTTTTCAAAT	GGGATTTCAA	CTCCGCCAAG	TATCAGCTTT
701	CTGCCTTCGG	TCAATTGGTG	GACATCGTAG	CCCTGTCCGA	TACGGATATT
751	CGTCATCGTT	TGTGTTCTCTG	A		

This encodes a protein having amino acid sequence <SEQ ID 798>:

1	MSYRASSSPD	FLEVETAPLI	FLPLLPKASM	KKLMVEPVPM	PMYSFSGTNS
51	TAFSAAMRLS	SSSCVVIFLSF	GKPYQQTAAI	LTFECTSWPP	RSNPYQQYRR
101	LRLYAFHPPE	IAEFFVGFAF	DIDARNIDTQ	IGGDVGTHLR	NVRCEFGFLC
151	NHGRIDIDL	PTLRLNALIR	RTQKDAAVRI	FELCGGVGKM	AADVAQTCRT
201	EQRVGNVQQ	RVGIRMPEQP	FFKWFDFNSAK	YQLSAFGQLV	DIVALSDDI
251	RHRLCS*				

ORF122ng and ORF122-1 show 92.6% identity in 256 aa overlap:

		10	20	30	40	50	60
	orf122-1.pep	ISYWSSSPDFLEVDTAPLIFLPLLPKASMKKLMVEPVMPPIYSFSGTNSTAFSAAMRLS					
	orf122ng	MSYRASSSPDFLEVETAPLIFLPLLPKASMKKLMVEPVMPMYSFSGTNSTAFSAAMRLS					
45		10	20	30	40	50	60
	orf122-1.pep	SSSCVVIFLSFGKPYQQTAAILTFFECTSCPPRSNAYQQYRRLRLYAFHPPEIAEFFVGFAF					
	orf122ng	SSSCVVIFLSFGKPYQQTAAILTFFECTSWPPRSNPNYQQYRRLRLYAFHPPEIAEFFVGFAF					
50		70	80	90	100	110	120
	orf122-1.pep	DVDARNVYAQIGGDVGTHLRNVRREFGFLCNHGRIDIDRLPTLRLNALIRRTQKDAAVRI					
	orf122ng	DIDARNIDTQIGGDVGTHLRNVRCEFGFLCNHGRIDIDHLPRLNALIRRTQKDAAVRI					
55		130	140	150	160	170	180
	orf122-1.pep	FELCGGVGEMAADIAQTCRTEQRVGNVQQRIGIGVSEQPFFKWFDFNSAKYQLSAFGQLV					
	orf122ng	FELCGGVGKMAADVAQTCRTEQRVGNVQQRVGIRMPEQPFFKWFDFNSAKYQLSAFGQLV					
60		190	200	210	220	230	240

5 orf122-1.pep 250  
 orf122ng 250  
 DIVALSDTDVHRRLCSX  
 |||||:|||||  
 DIVALSDTDIRHRLCSX

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 95

10 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 799>:

1 ..GCCGGCGCGA GTGCGAACAA CATTTCGCGG CGTTTTCGCG AAACACCCGT  
 51 CGCTGTCAGC GTTACCCTGA TCGGCACGGT ACTTGCCGTC ATGCTGCCCG  
 101 TTACCGAATA TGAAAACCTC CTGCTGCTTA TCGGCTCGGT ATTTGCGCCG  
 151 ATGGGGCGGA TTTTGATTGC CGACTTTTTC GTCTTGAAAC GCGGTGA

15 This corresponds to the amino acid sequence <SEQ ID 800; ORF125>:

1 ..AGASANNISA RFAETPVAVS VTLLGTVLAV MLPVTEYENF LLLIGSVFAP  
 51 MGGFDCRLFR LETA\*

Further work revealed the complete nucleotide sequence <SEQ ID 801>:

1 ATGTGCGGCA ATGCCTCCTC TCCTTCATCT TCCTCCGCCA TCGGGCTGAT  
 20 51 TTGGTTCGGC GCGGCGGTAT CGATTGCCGA AATCAGCACG GGTACGCTGC  
 101 TTGCGCCTTT GGGCTGGCAG CGCGGTCTGG CGGCTCTACT TTTGGGTCAT  
 151 GCCGTCGGCG GCGCGCTGTT TTTGCGGCG GCGTATATCG GCGCACTGAC  
 201 CCGACGCAGC TCATGGAAA GCGTGCGCCT GTCGTTCCGG AAACGCGGTT  
 251 CAGTGCTGTT TTCCGTGGCG AATATGCTGC AACTGGCCGG CTGGACGCGC  
 25 301 GTGATGATTT ACGCCGGCGC AACGGTCAGC TCCGCTTTGG GCAAAGTGTT  
 351 GTGGGACGGC GAATCTTTTG TCTGGTGGGC ATTGGCAAAC GCGCGCTGA  
 401 TTGTGCTGTG GCTGGTTTTT GCGCACGCA AAACAGGCGG GCTGAAAACC  
 451 GTTTCGATGC TGCTGATGCT GTTGGCGGTT CTGTGGCTGA GTGCCGAAGT  
 501 CTTTCCACG GCAGGCAGCA CCGCCGACA GGTTCAGAC GGCATGAGTT  
 30 551 TCGGAACGGC AGTCGAGCTG TCCGCCGTA TGCCGCTTTC CTGGCTGCCG  
 601 CTTGCCGCGC ACTACACGCG CCACGCGCGC CGCCCGTTTG CGGCAACCT  
 651 GACGGCAACG CTCGCTACA CGCTGACCG CTGCTGGATG TATGCCTTGG  
 701 GTTTGGCAGC GCGGTTGTT ACCGGAGAAA CCGACGTGGC AAAAATCCTG  
 751 CTGGGCGCAG GTTGGGTGTC GGCAGGCATT TTGGCGGTGC TCCTCTCCAC  
 35 801 CGTTACCACA ACGTTTCTCG ATGCCTATTC CGCCGGCGCG AGTGCGAACA  
 851 ACATTTCCGC GCGTTTTCG GAAACACCG TCGCTGTCGG CGTTACCTCG  
 901 ATCGGCACGG TACTTGCCGT CATGCTGCC GTTACCGAAT ATGAAAACCT  
 951 CCTGCTGCTT ATCGGCTCGG TATTTGCGCC GATGGCGCGG GTTTTGATTG  
 40 1001 CCGACTTTTT CGTCTGAAA CCGCGTGAGG AGATTGAAG CTTTGACTTT  
 1051 GCCGGACTGG TTCTGTGGCT TCGGGGCTTC ATCCTCTACC GCTTCTGCT  
 1101 CTCGTCCGCG TGGGAAAGCA GCATCGGTCT GACCGCCCC GTAATGTCTG  
 1151 CCGTTGCCAT TGCCACCGTA TCGGTACGCC TTTTCTTTAA AAAAACCCTA  
 1201 TCTTTACAAA GGAACCCGTC ATGA

This corresponds to the amino acid sequence <SEQ ID 802; ORF125-1>:

45 1 MSGNASSPSS SSAIGLIWFG AAVSIAEIST GTLLAPLGWQ RGLAALLLGH  
 51 AVGGALFFAA AYIGALTGRS SMESVRLSFG KRGSVLFSA NMLQLAGWTA  
 101 VMIYAGATVS SALGKVLWDG ESFVWWALAN GALIVLWLVF GARKTGGLKT  
 151 VSMLLMLLAV LWLSAEVFST AGSTAAQVSD GMSFGTAVEL SAVMPLSWLP  
 201 LAADYTRHAR RPFATLTAT LAYLTGCMW YALGLAAALF TGETDVAKIL  
 50 251 LGAGLGAAGI LAVVLSTVTT TFLDAYSAGA SANNISAREFA ETPVAVGVTL  
 301 IGTVLAVMLP VTEYENFLLL IGSVFAPMAA VLIADFFVLK RREEIEGFDF  
 351 AGLVLWLAGF ILYREFLLSSG WESSIGLTAP VMSAVAIATV SVRLFFKKTQ  
 401 SLQRNFS\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF125 shows 76.5% identity over a 51aa overlap with an ORF (ORF125a) from strain A of *N. meningitidis*:

```

5      orf125.pep      AGASANNISARFAETPVAVSVTLIGTVLAV
      orf125a      KILLGAGLGAAGILAVVLSTVTTTFLDAYSAGVSAANNISAKLSEIPIAVAVAVVGTLLAV
                  250      260      270      280      290      300

10     orf125.pep      MLPVTEYENFLLLLIGSVFAPMGGFDCRLFRLETAX
      orf125a      LLPVTEYENFLLLLIGSVFAPMAAVLIADFFVLKRREEIEG
                  310      320      330      340

```

15 The ORF125a partial nucleotide sequence <SEQ ID 803> is:

```

1      ATGTCGGGCA ATGCCTCCTC TCNTTCATCT TCCGCCGCCA TCGGGCTGAT
51     TTGGTTCGGC GCGGCGGTAT CGATTGCCGA AATCAGCACG GGTACACTGC
101    TTGCGCCTTT GGGCTGGCAG CGCGGTCTGG CNGCTCTGCT TTTGGGTCAT
151    GCCGTCGGCG GCGCGCTGTT TTTTGGCGCG GCGTATATCG GCGCACTGAC
201    CCGACNCANC TCGATGGAAG GCGTGCCTGT GTCGTTCGGC AAACGCGGTT
251    CAGTGCTGTT TTCCGTGGCG AATATGCTGC AACTGGCCGG CTGGACGGCG
301    GTGATGATTT ACGCCGCGCG AACGGTCAGC TCCGCTTTGG GCAAAGTGTT
351    GTGGGACGGC GAATCTTTTG TCTGGTGGGC ATTGGCAAAC GGCAGCGCTGA
401    TTGTGCTGTG GCTGGTTTTC GGCGCACGCA AAACAGGCGG GCTGAAAACC
451    GTTTCGATGC TGCTGATGCT GTTGGCGGTT CTGTGGCTGA GTGCCGAANT
501    NTTTTCCACG GCAGGCAGCA CCGCCGCANN GGTNNCAGAC GGCATGAGTT
551    TCGGAACGGC AGTCGAGCTG TCCGCCGTNA TGCCGCTTTC TTGGCTGCCG
601    CTGGCCGCCG ACTACACGCG CCACGCGCGC CGCCCGTTTG CGGCAACCTT
651    GACGGCAACG CTGCGCTACA CGCTGACCGG CTGTGGATG TATGCCTTGG
701    GTTTGGCAGC GCGGTTGTTC ACCGGAGAAA CCGACGTGGC AAAAATCCTG
751    CTGGGCGCAG GTTTGGGTGC GGCAGGCATT TTGGCGGTGC TCCTGTGCAG
801    CGTTACCACC GTTTTCTCG ATGCNTACTC CGCCGGCGTA AGTGCCAACA
851    ATATTCCCGC CAAACTTTCG GAAATACCNA TCGCGGTTGC CGTCGCCGTT
901    GTCGGCACAC TGCTTGCCGT CCTCCTGCCC GTTACCGAAT ATGAAAACCT
951    CCTGCTGCTT ATCGGCTCGG TATTGCGGCC GATGGCGGCG GTTTTGATTG
1001   CCGACTTTTT CGTCTTGAAA CGGCGTGAGG AGATTGAAGG C..

```

This encodes a protein having the partial amino acid sequence <SEQ ID 804>:

```

1      MSGNASSXSS SAAIGLIWFG AAVSIAEIST GTLLAPLGWQ RGLAALLLGH
51     AVGGALFFAA AYIGALTGXX SMESVRLSFG KRGSVLFSSA NMLQLAGWTA
101    VMIYAGATVS SALGKVLWDG ESFVWWALAN GALIVLWLVF GARKTGGLKT
151    VSMMLMLLAV LWLSAEXFST AGSTAAXVXD GMSFGTAVEL SAVMPLSWLP
201    LAADYTRHAR RPFATLTAT LAYTLTGCMW YALGLAAALF TGETDVAKIL
251    LGAGLGAAGI LAVVLSTVTT TFLDAYSAGV SANNISAKLS EIPIAVAVAV
301    VGTLLAVLLP VTEYENFLLL IGSVFAPMAA VLIADFFVLK RREEIEG..

```

45 ORF125a and ORF125-1 show 94.5% identity in 347 aa overlap:

```

      orf125a.pep      MSGNASSXSSSAAIGLIWFGAAVSIAEISTGTLLAPLGWQ RGLAALLLGHAVGGALFFAA
      orf125-1      MSGNASSPSSSSAIGLIWFGAAVSIAEISTGTLLAPLGWQ RGLAALLLGHAVGGALFFAA
                  10      20      30      40      50      60

50     orf125a.pep      AYIGALTGXXSMESVRLSFGKRGSVLFSSVANMLQLAGWTAVMIYAGATVSSALGKVLWDG
      orf125-1      AYIGALTGRSSMESVRLSFGKRGSVLFSSVANMLQLAGWTAVMIYAGATVSSALGKVLWDG
                  70      80      90      100      110      120

55     orf125a.pep      ESFVWWALANGALIVLWLVFGARKTGGLKTVSMLMLLAVLWLSAEXFSTAGSTAAXVXD
      orf125-1      ESFVWWALANGALIVLWLVFGARKTGGLKTVSMLMLLAVLWLSAEXFSTAGSTAAXVXD
                  130      140      150      160      170      180

60     orf125a.pep

```

	orfl25-1	ESFVWWALANGALIVLWLVFGARKTGGLKTVSMMLMLLAVLWLSAEVFSTAGSTAAQVSD	130	140	150	160	170	180
5	orfl25a.pep	GMSFGTAVELSAVMPLSWLPLAADYTRHARRPFAATLTATLAYTLTGCWMYALGLAAALF	190	200	210	220	230	240
10	orfl25-1	GMSFGTAVELSAVMPLSWLPLAADYTRHARRPFAATLTATLAYTLTGCWMYALGLAAALF	190	200	210	220	230	240
	orfl25a.pep	TGETDVAKILLGAGLGAAGILAVVLSTVTTTFLDAYSAGVSANNISAKLSEIPIAVAVAV	250	260	270	280	290	300
15	orfl25-1	TGETDVAKILLGAGLGAAGILAVVLSTVTTTFLDAYSAGVSANNISARFAETPVAVGVTL	250	260	270	280	290	300
	orfl25a.pep	VGTLAVLLPVTEYENFLLIGSVFAPMAAVLIADFFVLKRREEIEG	310	320	330	340		
20	orfl25-1	IGTVLAVMLPVTEYENFLLIGSVFAPMAAVLIADFFVLKRREEIEGDFDFAGLVWLWLAGF	310	320	330	340	350	360

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF125 shows 86.2% identity over a 65aa overlap with a predicted ORF (ORF125ng) from

25 *N.gonorrhoeae*:

	orfl25.pep	AGASANNISARFAETPVAVSVTLIGTVLAV	30
	orfl25ng	KILLGAGLGITGILAVVLSTVTTTFLDYSAGASANNISARFAEIPVAVGVTLIRTVLAV	308
30	orfl25.pep	MLPVTEYENFLLIGSVFAPM-GGFCRLFRLETA	64
	orfl25ng	MLPVTEYKNFLLIRSVFGPMAGGFCRLFLCKTA	343

An ORF125ng nucleotide sequence <SEQ ID 805> was predicted to encode a protein having amino acid sequence <SEQ ID 806>:

35	1	MSGNASSPSS	SAAIGLVWFG	AAVSIAEIST	GTLLAPLGWQ	RGLAALLLGH
	51	AVGGALFFAA	AYIGALTGRS	SMESVRLSFG	KCGSVLFSVA	NMLQLAGWTA
	101	VMIYVGATVS	SALGKVLWDG	ESFVWWALAN	GALIVLWLVF	GARRTGGLKT
	151	VSMMLMLLAV	LWLSVEVFAS	SGTNAAPAVS	DGMTFGTAVE	LSAVMPLSWL
40	201	PLAADYTRQA	RRPFAATLTA	TLAYTLTGCW	MYALGLAAAL	FTGETDVAKI
	251	LLGAGLGITG	ILAVVLSTVT	TTFLDYSAG	ASANNISARF	AEIPVAVGVT
	301	LIRTVLAVML	PVTEYKNFLL	LIRSVFGPMA	GGFCRLFLCL	KTA*

Further work revealed the following gonococcal DNA sequence <SEQ ID 807>:

	1	ATGTCGGGCA	ATGCCTCCTC	TCCTTCATCT	TCCGCCGCCA	TCGGGCTGGT
45	51	TTGGTTCGGC	GCGGCGGTAT	CGATTGCCGA	AATCAGCACG	GGTACGCTGC
	101	TCGCCCCCTT	GGGCTGGCAG	CGCGGTCTGG	CGGCCCTGCT	TTTGGGTCAT
	151	GCCGTCGGCG	GCGCGCTGTT	TTTTCGGCGG	GCGTATATCG	GCGCACTGAC
	201	CGGACGCAGC	TCGATGAAA	GTGTGCGCCT	GTCGTTCCGG	AAATGCGGTT
	251	CAGTGCTGTT	TTCCGTGGCG	AATATGCTGC	AACTGGCCGG	CTGGACGGCG
	301	GTGATGATTT	ACGTCCGGCG	AACGGTCAGC	TCCGCTTTGG	GCAAAGTGTT
50	351	GTGGGACGGC	GAATCCTTTG	TCTGGTGGGC	ATTGGCAAAC	GGCGCACTGA
	401	TCGTGCTGTG	GCTGGTTTTC	GGCGCACGCA	GAACGGGCGG	GCTGAAAACC
	451	GTTTCGATGC	TGCTGATGCT	GCTTGCCGTG	TTGTGGTTGA	GCGTCGAAGT
	501	GTTTCGCTTCG	TCCGGCACAA	ACGCCGCGCC	CGCCGTTTCA	GACGGCATGA
	551	CCTTCGGAAC	GGCAGTCGAA	CTGTCCGCGG	TCATGCCGCT	TTCTGGCTG
55	601	CCGCTGGCCG	CGACTACAC	GCGCCAAGCA	CGCCGCCCGT	TTGCGGCAAC
	651	CCTGACGGCA	ACGCTCGCCT	ATACGCTGAC	GGGTGCTGG	ATGTATGCCT
	701	TGGGTTTGGC	GGCGGCTCTG	TTTACCGGAG	AAACCGACGT	GGCGAAAATC
	751	CTGTTGGGCG	CGGGCTTGGG	CATAACGGGC	ATTCTGGCAG	TCGCTCTCTC
	801	CACCGTTACC	ACAACGTTTC	TCGATACCTA	TTCCGCCGGC	GCGAGTGCGA
60	851	ACAACATTTT	CGCGCGTTT	GCGGAAATAC	CCGTGCTGT	CGGCGTTACC

5  
 901 CTGATCGGCA CGGTGCTTGC CGTCATGCTG CCCGTTACCG AATATAAAAA  
 951 CTTCCTGCTG CTTATCGGCT CGGTATTTCG GCCGATGGCG GCGGTTTTGA  
 1001 TTGCCGACTT TTTCGTCTTA AAACGGCGTG AGGAGATTGA AGGCTTTGAC  
 1051 TTTGCCGGAC TGGTTCTGTG GCTGGCAGGC TTCATCCTCT ACCGCTTCCT  
 1101 GCTCTCGTCC GGTGGGAAA GCAGCATCGG TCTGACCGCC CCCGTAATGT  
 1151 CTGCCGTTGC CATTGCCACC GTATCGGTAC GCCTTTTCTT TAAAAAACC  
 1201 CAATCTTTAC AAAGAACCC GTCATGA

This corresponds to the amino acid sequence <SEQ ID 808; ORF125ng-1>:

10  
 1 MSGNASSPSS SAAIGLVWFG AAVSIAEIST GTLLAPLGWQ RGLAALLLGH  
 51 AVGGALFFAA AYIGALTGRS SMESVRLSFG KCGSVLFSVA NMLQLAGWTA  
 101 VMIYVGATVS SALGKVLWDG ESFVWWALAN GALIVLWLVF GARRTGGLKT  
 151 VSMLLMLLAV LWLSVEVFAS SGTNAAPAVS DGMTFGTAVE LSAVMPLSWL  
 201 PLAADYTRQA RRPFAATLTA TLAYTLTGCW MYALGLAAL FTGETDVAKI  
 251 LLGAGLGITG ILAVVLSTVT TTFLDTYSAG ASANNISARF AEIPVAVGVT  
 15  
 301 LIGTVLAVML PVTEYKNFLL LIGSVFAPMA AVLIADFFVL KRREEIEGFD  
 351 FAGLVLWLAG FILYRFLSS GWESSIGLTA PVMSAVAIAT VSVRLFFKKT  
 401 QSLQRNPS\*

ORF125ng-1 and ORF125-1 show 95.1% identity in 408 aa overlap:

20  
 orf125-1.pep 10 20 30 40 50 60  
 MSGNASSPSSSSAIGLIWFGAAVSIAEISTGTLLAPLGWQ RGLAALLLGHAVGGALFFAA  
 orf125ng-1 10 20 30 40 50 60  
 MSGNASSPSSSSAIGLVWFGAAVSIAEISTGTLLAPLGWQ RGLAALLLGHAVGGALFFAA  
 25  
 orf125-1.pep 70 80 90 100 110 120  
 AYIGALTGRSSMESVRLSFGKRGSVLFSVANMLQLAGWTAVMIYAGATVSSALGKVLWDG  
 orf125ng-1 70 80 90 100 110 120  
 AYIGALTGRSSMESVRLSFGKCGSVLFSVANMLQLAGWTAVMIYVGATVSSALGKVLWDG  
 30  
 orf125-1.pep 130 140 150 160 170 179  
 ESFVWWALANGALIVLWLVFGARKTGGLKTVSMLLMLLAVLWLSAEVFASTAGSTAAQ-VS  
 orf125ng-1 130 140 150 160 170 180  
 ESFVWWALANGALIVLWLVFGARRTGGLKTVSMLLMLLAVLWLSVEVFASSTNAAPAVS  
 35  
 orf125-1.pep 180 190 200 210 220 230 239  
 DGMSFGTAVELSAVMPLSWLPLAADYTRHARRPFAATLTATLAYTLTGCW MYALGLAAL  
 orf125ng-1 180 190 200 210 220 230 240  
 DGMTFGTAVELSAVMPLSWLPLAADYTRQARRPFAATLTATLAYTLTGCW MYALGLAAL  
 40  
 orf125-1.pep 240 250 260 270 280 290 299  
 FTGETDVAKILLGAGLGAAGILAVVLSTVTTTFLDAYASAGASANNISARFAETPVAVGVT  
 orf125ng-1 240 250 260 270 280 290 300  
 FTGETDVAKILLGAGLGITGILAVVLSTVTTTFLDTYSAGASANNISARFAEIPVAVGVT  
 45  
 orf125-1.pep 300 310 320 330 340 350 359  
 LIGTVLAVMLPVTEYENFLLIGSVFAPMAAVLIADFFVLKRREEIEGFD FAGLVLWLAG  
 orf125ng-1 300 310 320 330 340 350 360  
 LIGTVLAVMLPVTEYKNFLLIGSVFAPMAAVLIADFFVLKRREEIEGFD FAGLVLWLAG  
 50  
 orf125-1.pep 360 370 380 390 400  
 FILYRFLSSGWESSIGLTAPVMSAVAIATVSVRLFFKKTQSLQRNPSX  
 orf125ng-1 360 370 380 390 400  
 FILYRFLSSGWESSIGLTAPVMSAVAIATVSVRLFFKKTQSLQRNPSX

60 Based on this analysis, including the presence of putative leader sequence and transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and



*N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 96

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 809>:

```

5       1  ATGACCCGTA TCGCCATCCT CGGCGGCGGC CTCTCGGGAA GGCTGACCGC
      51  GTTGACAGTT GCAGAACAAAG GTTATCAGAT TGCACCTTTC GATAAAAGCT
     101  GCCGCCGGGG CGAACACGCC GCGCCTATG TAGCCGCGCG CATGCTCGCG
     151  CCTGCAGCGG A.ACGGTCGA AGCCACGCCC GAAGTGGTCA GGCTGGGCAG
     201  GCAGAGCATC CCGCTTTGGC GCGGCATCCG ATGCCGTCTG AACACGCACA
    110  251  CGATGATGCA GGA AACCGGC AGCCTGATTG TATGGCACGG GCAGGACAAG
      301  CCATTATCCA GCGAGTTCGT CCGCCATCTC AAACGCGGCG GCGT.ACGBA
      351  TGACGAAATC GTCCGTTGGC GCGCCGACGA CATCGCCGAA CGCGAACCGC
      401  AACTCGGCGG ACGTTTTTAA GACGGCATCT ACCTGCCGAC CGAAGC.CAG
      451  CTCGACGGGC GGCAATTATA GTCTGCACTT GCCGACGCTT TGGACGAACT
    115  501  GAACGTCCCC TGCCATTGGG AACACGAATG CGTCCCCGAA GCCTGCAAG..
  
```

This corresponds to the amino acid sequence <SEQ ID 810; ORF126>:

```

      1  MTRIAILGGG LSGRLTALQL AEQGYQIALF DKSCRRGEHA AAYVAAAML A
     51  PAAXTVEATP EVVRLGRQSI PLWRGIRCRL NHTMMQENG SLIVWHGQDK
    101  PLSSEFVRHL KRGXITDDEI VRWRADDIAE REPQLGGREF DGIYLPTEXQ
    151  LDGRQLXSAL ADALDELNVP CHWEHECVPE ACK...
  
```

Further work revealed the complete nucleotide sequence <SEQ ID 811>:

```

      1  ATGACCCGTA TCGCCATCCT CGGCGGCGGC CTCTCGGGAA GGCTGACCGC
     51  GTTGACAGTT GCAGAACAAAG GTTATCAGAT TGCACCTTTC GATAAAGGCT
    101  GCCGCCGGGG CGAACACGCC GCGCCTATG TTGCCGCCGC CATGCTCGCG
    151  CCTGCAGCGG AAGCGGTCGA AGCCACGCCC GAAGTGGTCA GGCTGGGCAG
     201  GCAGAGCATC CCGCTTTGGC GCGGCATCCG ATGCCGTCTG AACACGCACA
     251  CGATGATGCA GGA AACCGGC AGCCTGATTG TGTGGCACGG GCAGGACAAG
     301  CCATTATCCA GCGAGTTCGT CCGCCATCTC AAACGCGGCG GCGTAGCGGA
     351  TGACGAAATC GTCCGTTGGC GCGCCGACGA CATCGCCGAA CGCGAACCGC
     401  AACTCGGCGG ACGTTTTTCA GACGGCATCT ACCTGCCGAC CGAAGGCCAG
     451  CTCGACGGGC GGCAATATT GTCTGCACTT GCCGACGCTT TGGACGAACT
     501  GAACGTCCCC TGCCATTGGG AACACGAATG CGTCCCCGAA GGCTGCAAG
     551  CCAATACGA CTGGCTGATC GACTGCCGCG GCTACGGCGC AAAAACCGCG
     601  TGGAACCAAT CCCCCGAGCA CACCAGCACC CTGCGCGGCA TACGCGGCGA
    115  651  AGTGGCGCGG GTTTACACAC CCGAAATCAC GCTCAACCGC CCCGTGCGTC
      701  TGCTCCATCC GCGTTATCCG CTCTACATCG CCCCAGAAAG AAACACGTC
      751  TTCGTATCCG GCGCGACCCA AATCGAAAGC GAAAGCCAAG CCCCCGCCAG
      801  CGTGCCTTCA GGGTTGGAAC TCTGTCCGC ACTCTATGCC ATCCACCCCG
      851  CTTTCGCGCA AGCCGACATC CTCGAAATCG CCACCGGCCT GCGCCCCACG
    140  901  CTCAACCAAC ACAACCCCGA AATCCGTTAC AACCGCGCCC GACGCTGAT
      951  TGAAATCAAC GGCCTTTTCC GCCACGGTTT CATGATCTCC CCCGCCGTAA
    1001  CCGCCGCGC CGCCAGATTG GCAGTGGCAC TGTTTGACGG AAAAGACGCG
    1051  CCCGAACGCG ATAAAGAAAG CGGTTTGGCG TATATCCGAA GACAAGATTA
    1101  A
  
```

45 This corresponds to the amino acid sequence <SEQ ID 812; ORF126-1>:

```

      1  MTRIAILGGG LSGRLTALQL AEQGYQIALF DKGCRRGEHA AAYVAAAML A
     51  PAEEAVEATP EVVRLGRQSI PLWRGIRCRL NHTMMQENG SLIVWHGQDK
    101  PLSSEFVRHL KRGGVADDEI VRWRADDIAE REPQLGGREFS DGIYLPTEGQ
    151  LDGRQILSAL ADALDELNVP CHWEHECVPE GLQAQYDWLI DCRGYGAKTA
    201  WNQSPEHTST LRGIQGEVAR VYTPFITLNR PVRLHPRYP LYIAPKENHV
    251  FVIGATQIES ESQAPASVRS GLELLSALYA IHPAFGEADI LEIATGLRPT
    301  LNHHNPEIRY NRARRLIEIN GLFRHGFMS PAVTAAARL AVALFDGKDA
    351  PERDKESGLA YIRRQD*
  
```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF126 shows 90.0% identity over a 180aa overlap with an ORF (ORF126a) from strain A of *N.*

*meningitidis*:

```

5      orf126.pep  MTRIAILGGGLSGRLTALQLAEQGYQIALFDKSCRRGEHAAAYVAAAMLAPAAXTVEATP
      orf126a     MTRIAILGGGLSGRLTALQLAEQGYQIALFDKGCRRGEHAAAYVAAAMLAPAAEAVEATP
10      orf126.pep  EVVRLGRQSIPLWRGIRCLNTHMTMMQENGLIVVHGQDKPLSSEFVRHLKRGXDTDEI
      orf126a     EVVRLGRQXIPLWRGIRCHLKT PAMMXENGLIVVHGQDKPLSNEFVRHLKRGVADDXI
15      orf126.pep  VRWRADDIAEREPQLGGRFXDGIYLPTEXQLDGRQLXSALADALDELNPCHWEHECVPE
      orf126a     VRWRADDIAEREPQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNPCHWEHECAPE
20

```

The complete length ORF126a nucleotide sequence <SEQ ID 813> is:

```

1  ATGACCCGTA TCGCCATCCT CGGCGGCGGC CTCTCNGGAA GGCTGACCGC
51 ACTGCAGCTT GCAGAACAAAG GTTATCAGAT TGCACTTTTC GATAAAGGCT
101 GCCGCCGGGG CGAACACGCC GCGCCTATG TTGCCGCCGC CATGCTCGCG
25 151 CCTGCGGCGG AAGCGGTCGA AGCCACGCCT GAAGTGGTCA GGCTGGGCAG
201 GCAGANCATC CCGCTTTGGC GCGGCATCCG ATGCCATCTG AAAACGCCTG
251 CCATGATGCA NGAAAACGGC AGCCTGATTG TGTGGCACGG GCAGGACAAA
301 CCTTTATCCA ACGAGTTCGT CCGCATCTC AAACGCGGCG GCGTAGCGGA
351 TGACNAAATC GTCCGTGGC GCGCCGACGA CATCGCCGAA CGCGAACC GC
401 AACTCGGCGG ACGTTTTTCA GACGGCATCT ACCTGCCGAC CGAAGGCCAG
451 CTCGACGGGC GGCAAAATATT GTCTGCACTT GCCGACGCTT TGGACGAAC
501 GAACGTCCCC TGCCATTGGG AACACGAATG TGCCCCGAA GACTTGCAAG
551 CCCAATACGA CTGGCTGATC GACTGCCGCG GCTACGGCGC AAAAACCGCG
601 TGGAACCAAT CCCCCGANNA NACCAGCACC CTGCGCGGCA TACGCGGCGA
35 651 AGTGGCGCGG GTTTACACAC CCGAAATCAC GCTCAACCGC CCCGTGCGCC
701 TGCTACACCC GCGCTATCCG CTNTACATCG CCCCAGAAAG AAACCNCGTC
751 TTCGTCATCG GCGCGACCCA AATCGAAAGC GAAAGCCAAG CACCTGCCAG
801 CGTGGCTTCC GGGCTGGAAC TCTTATCCGC ACTCTATGCC GTCCACCCCG
851 CCTTCGGCGA AGCCGACATC CTCGAAATCG CCACCGGCCT GCGCCCCACG
40 901 CTCAATCACC ACAACCCGAA AATCCGTTAC AACCGCGCCC GACGCCTGAT
951 TGAAATCAAC GGCCTTTTCC GCCACGGTTT CATGATCTCC CCCGCCGTAA
1001 CCGCCGCCGC CGTCAGATTG GCAGTGGCAC TGTTTGACGG AAAAGANGCG
1051 CCCGAACGCG ATGAAGAAAG CGGTTTGGCG TATATCCGAA GACAAGATTA
1101 A

```

45 This encodes a protein having amino acid sequence <SEQ ID 814>:

```

1  MTRIAILGGG LSGRLTALQL AEQGYQIALF DKGCRREGEH AAYVAAAMLA
51  PAAEAVEATP EVVRLGRQXI PLWRGIRCHL KTPAMMXENG SLIVVHGQDK
101 PLSNEFVRHL KRGGVADDXI VRWRADDIAE REPQLGGRFS DGIYLPTEGQ
50 151 LDGRQILSAL ADALDELNPV CHWEHECAPE DLQAQYDWLI DCRGYGAKTA
201 WNQSPXXTST LRGIRGEVAR VYTPEITLNR PVRLHPRYP LYIAPKENXV
251 FVIGATQIES ESQAPASVRS GLELLSALYA VHPAFGEADI LEIATGLRPT
301 LNHNNPEIRY NRARRLIEIN GLFRHGFMI S PAVTAAAVRL AVALFDGKXA
351 PERDEESGLA YIRRQD*

```

ORF126a and ORF126-1 show 95.4% identity in 366 aa overlap:

```

55      orf126a.pep  MTRIAILGGGLSGRLTALQLAEQGYQIALFDKGCRRGEHAAAYVAAAMLAPAAEAVEATP
      orf126-1     MTRIAILGGGLSGRLTALQLAEQGYQIALFDKGCRRGEHAAAYVAAAMLAPAAEAVEATP
60

```

		70	80	90	100	110	120
	orf126a.pep	EVVRLGRQXIPLWRGIRCHLKT	PAMMXENGSLIVWHGQDKPLS	NEFVRHLKRGGVADDDXI			
	orf126-1	EVVRLGRQSIPLWRGIRCLNTH	TMMQENGSLIVWHGQDKPLS	SSEFVRHLKRGGVADDEI			
5		70	80	90	100	110	120
	orf126a.pep	VRWRADDIAEREPQLGGRFSD	GIYLPTEGQLDGRQILSALA	DALDELNVPCHEHECAPE			
10	orf126-1	VRWRADDIAEREPQLGGRFSD	GIYLPTEGQLDGRQILSALA	DALDELNVPCHEHECVPE			
		130	140	150	160	170	180
	orf126a.pep	DLQAQYDWLIDCRGYGAKTAW	NQSPXXTSTLRGIRGEVARV	YTPFITLNRPVRLHPRYP			
15	orf126-1	GLQAQYDWLIDCRGYGAKTAW	NQSPEHTSTLRGIRGEVARV	YTPFITLNRPVRLHPRYP			
		190	200	210	220	230	240
	orf126a.pep	LYIAPKENXVFIGATQIESES	QAPASVRSGLLELLSALYA	VHPAFGEADILEIATGLRPT			
20	orf126-1	LYIAPKENHVFIGATQIESES	QAPASVRSGLLELLSALYA	IHPAFGEADILEIATGLRPT			
		250	260	270	280	290	300
	orf126a.pep	LNHHNPEIRYNRARRLIEIN	GLFRHGFMISSPAVTA	AAVRLAVALFDGKXAPER	DEESGLA		
25	orf126-1	LNHHNPEIRYNRARRLIEIN	GLFRHGFMISSPAVTA	AAARLAVALFDGKDAPER	KESGLA		
		310	320	330	340	350	360
	orf126a.pep	YIRRQDX					
30	orf126-1	YIRRQDX					

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF126 shows 90% identity over a 180 aa overlap with a predicted ORF (ORF126ng) from

*N.gonorrhoeae*:

40	orf126.pep	MTRIAVLGGGLSGRLTALQLAEQGYQIALFDKSCRRGEHAAAYVAAAMLAPAAXTVEATP	60
	orf126ng	MTRIAVLGGGLSGRLTALQLAEQGYQIELFDKGTQRGEHAAAYVAAAMLAPAAEVEATP	60
	orf126.pep	EVVRLGRQSIPLWRGIRCLNTH	
45	orf126ng	EVIRLGRQSIPLWRGIRCLNTH	
	orf126.pep	VRWRADDIAEREPQLGGRFXDGIYLP	
	orf126ng	VRWRADEIAEREPQLGGRFSDGIYLP	

50 An ORF126ng nucleotide sequence <SEQ ID 815> was predicted to encode a protein having amino acid sequence <SEQ ID 816>:

	1	MTRIAVLGGGLSGRLTALQLAEQGYQIELFDKGTQRGEHAAAYVAAAMLAPAAEVEATP
	51	EVIRLGRQSIPLWRGIRCLNTH
	101	PLSSEFVRHLKRGGVADDEI
55	151	LDGRQILSALADALDELNVPCHEHECAPE
	201	WNQSPEHTSTLRGIRGEVARV
	251	SSSARPKSKAKPPPAYVP
	301	LNHHNPEIRYSRERRLIEIN
	351	PERDEESGLAYIGRQD*

60 Further work revealed the following gonococcal DNA sequence <SEQ ID 817>:

1 ATGACCCGTA TCGCCGTCCT CGGAGGCGGC CTTCCGGAA GGCTGACCGC  
51 ATTGCAGCTT GCAGAACAAAG GTTATCAGAT TGAACCTTTC GACAAGGGCA  
101 CCCGCCAAGG CGAACACGCC GCCGCCTATG TTGCCGCCGC GATGCTCGCG  
151 CCTGCGGCGG AAGCGGTGCA GGCAACGCCC GAAGTCATCA GGCTGGGCGAG  
5 201 GCAGAGCATT CCGCTTTGGC GCGGCATCCG ATGCCGTCTG AACACGCTCA  
251 CGATGATGCA GGAACACGGC AGCCTGATTG TGTGGCACGG GCAGGACAAG  
301 CCATTATCCA GCGAGTTCGT CCGCCATCTC AAACGCGGCG GCGTAGCGGA  
351 TGACGAAATC GTCCGTTGGC GCGCCGATGA AATCGCCGAA CGCGAACCGC  
401 AACTCGGCGG ACCTTTTTC GACGGCATCT ACCTGCCGAC CGAAGGCCAG  
10 451 CTCGACGGGC GGCAATATT GTCTGCACTT GCCGACGCTT TGGACGAACT  
501 GAACGTCCCT TGCCATTGGG AACACGAATG CGCCCCCAA GACCTGCAAG  
551 CCCAATACGA CTGGGTAATC GACTGCCGGG GCTACGGCGC GAAAACCGCG  
601 TGGAAACCAAT CCCCCGAGCA CACCAGCACC TTGCGCGGCA TACGCGGCGA  
651 AGTGGCGCGG GTTTACACGC CCGAAATCAC GCTCAACCGC CCCGTGCGCC  
15 701 TGCTGCACCC GCGCTATCCG CTCTACATCG CCCCAGAAAG AAACCACGTC  
751 TTCGTATCG GCGCGACCCA AATCGAAAGC GAAAGCCAAG CCCCCGCCAG  
801 CGTACGTTCC GGGCTGGAAC TCTTATCCGC GCTCTATGCC GTCCACCCCG  
851 CCTTCGGCGA AGCCGACATC CTCGAAATCG CCGCCGGCCT GCGCCCCACG  
901 CTCAACCACC ACAACCCCGA AATCCGCTAC AGCCGCGAAC GCCGCCTCAT  
20 951 CGAAATCAAC GGCCTTTTCC GGCACGGCTT TATGATTTCC CCGCCGTAA  
1001 CCGCCGCCGC CGTCAGATTG CGAGTGGCAC TGTTTGACGG AAAAGACGCG  
1051 CCCGAACGTG ATGAAGAAAG CGGTTTGGCG TATATCGGAA GACAAGATTA  
1101 A

This corresponds to the amino acid sequence <SEQ ID 818; ORF126ng-1>:

1 MTRIAVLGGG LSGRLTALQL AEQGYQIELF DKGTRQGEHA AAYVAAAMLA  
51 PAEEAVEATP EVIRLGRQSI PLWRGIRCL NTLTMMQENG SLIVWHGQDK  
101 PLSEFVRHL KRGGVADDEI VRWRADEIAE REPQLGGRFS DGIYLPTEGQ  
151 LDGRQILSAL ADALDELNVP CHWEHECAPQ DLQAQYDWVI DCRGYAKTA  
201 WNQSPEHTST LRGIRGEVAR VYTPEITLNR PVRLHPRYP LYIAPKENHV  
251 FVIGATQIES ESQAPASVRS GLELLSALYA VHPAFGEADI LEIAAGLRPT  
301 LNHHNPEIRY SRERRLIEIN GLFRHGFMI SPAVTAAAVRL AVALFDGKDA  
351 PERDEESGLA YIGRQD\*

ORF126ng-1 and ORF126-1 show 95.1% identity in 366 aa overlap:

35	orf126-1.pep	10	20	30	40	50	60
	orf126ng-1	10	20	30	40	50	60
40	orf126-1.pep	70	80	90	100	110	120
	orf126ng-1	70	80	90	100	110	120
45	orf126-1.pep	130	140	150	160	170	180
	orf126ng-1	130	140	150	160	170	180
50	orf126-1.pep	190	200	210	220	230	240
	orf126ng-1	190	200	210	220	230	240
55	orf126-1.pep	250	260	270	280	290	300
	orf126ng-1	250	260	270	280	290	300
60	orf126-1.pep	310	320	330	340	350	360
	orf126ng-1	310	320	330	340	350	360

orf126ng-1 LNHHNPEIRYSRERRLIEINGLFRHGFMI SPAVTA AAVRLAVALFDGKDAPERDEESGLA  
310 320 330 340 350 360

5 orf126-1.pep YIRRQDX  
|| ||||  
orf126ng-1 YIGRQDX

Furthermore, ORF126ng-1 shows homology to a putative *Rhizobium* oxidase flavoprotein:

10 gi|2627327 (AF004408) putative amino acid oxidase flavoprotein [Rhizobium etli]  
Length = 327  
Score = 169 bits (423), Expect = 3e-41  
Identities = 112/329 (34%), Positives = 163/329 (49%), Gaps = 25/329 (7%)  
15 Query: 3 RIAVLGGGLSGRLTALQLAEQGYQIELFDKGTRQGEHXXXXXXXXXXXXXXXXXXXXX 62  
RI V G G++G A QL G+++ L ++ G  
Sbjct: 2 RILVNGAGVAGLTVAWQLYRHGFRVTLAERAGTVGA-GASGFAGGMLAPWCERESAEPV 60  
Query: 63 IRLGRQSIPLWRGIRCLNLTMMQENGSLIVWHGQDKPLSSEFVRHLKRGVADDEIVR 122  
+ LGR + W + G+L+V G+D F R G DE+  
20 Sbjct: 61 LTLGRLAADWEAA-----LPGHVHRRGTLVVAGGRDTGELDRFSRRTS-GWEWLDEVA- 113  
Query: 123 WRADEIAEREPQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECAPQDL 182  
IA EP L GRF ++ E LD RQ L+ALA L++ + +  
25 Sbjct: 114 -----IAALEPDLAGRFRRLFFRQEAHLDPQALAAALAGLEDARMRLTLG---VVGES 165  
Query: 183 QAQYDWVIDCRGYGAKTAWNQSPEHTSTLRGIRGEVARVYTPEITLNRPVRLHPRYPLY 242  
+D V+DC G LRG+RGE+ V T E++L+RPVRLHPR+P+Y  
30 Sbjct: 166 DVDHDRVVDCTGAA-----QIGRLPGLRGVRGEMLCVETEVSLSRPVRLHPRHPIY 218  
Query: 243 IAPKENHVFFIGATQIESESQAPASVRSGLLELLSALYAVHPAFGEADILEIAAGLRPTLN 302  
I P++ + F++GAT IES+ P + RS +ELL+A YA+HPAFGEA + E AG+RP  
35 Sbjct: 219 IVPDRKNRFMVGATMIESDDGGPITARSLMELLNAAAYAMHPAFGEARVTETGAGVRPAYP 278  
Query: 303 HHNPEIRYSRERRLIEINGLFRHGFMI SP 331  
+ P R ++E R + +NGL+RHGF+++P  
Sbjct: 279 DNLP--RVTQEGRTLHVNGLYRHGFLLAP 305

This analysis suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 97

40 The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 819>:

1 ATGACTGATA ATCGGGGGTT TACGCTGGTT GAATTAATAT CAGTGGTCTT  
51 GATATTGTCT GTACTTGCTT TAATTGTTTA TCCGAGCTAT CGCAATTATG  
101 TTGAGAAAGC AAAGATAAAT GCAGTGCGGG CAGCCTTGTT AGAAAATGCA  
45 151 CATTTTATGG AAAAGTTTTA TCTGCAGAAAT GGGAGGTTTA AACAAACATC  
201 TACCAAGTGG CCAAGTTTGC CGATTAAAGA GGCAGAAGGC TTTGTATCC  
251 GTTTGAATGG AATCGtCGCG CGGG..GCTT TAGACAGTAA ATTCATGTTG  
301 AAGGCGGTAG CCATAGATAA AGATAAAAAT CCTTTTATTA TTAAGATGAA  
351 TGAAAATCTA GTAACCTTTA aTTTGCAAGA AGTCCGCCAG TTCGTGTAGT  
50 401 GACGGGCTGG ATTATTTTAA AGGAAATGAT AAGGACTGCA AGTTACTTAA  
451 GTAG

This corresponds to the amino acid sequence <SEQ ID 820; ORF127>:

1 MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN AVRAALLENA  
55 51 HFMEKFYLQN GRFKQTSTKW PSLPIKEAEG FCIRLNGIVA RXALDSKFML  
101 KAVAIDKDKN PFIKMNENL VTFICKKSAS SCSGGLDYFK GNDKDKLLK  
151 \*

Further work revealed the following DNA sequence <SEQ ID 821>:

```

      1  ATGACTGATA ATCGGGGGTT TACGCTGGTT GAATTAATAT CAGTGGTCTT
    51  GATATTGTCT GTACTTGCTT TAATTGTTTA TCCGAGCTAT CGCAATTATG
   101  TTGAGAAAGC AAAGATAAAT GCAGTGCGGG CAGCCTTGTT AGAAAATGCA
    151  CATTTTATGG AAAAGTTTTA TCTGCAGAAAT GGGAGGTTA AACAAACATC
    201  TACCAAGTGG CCAAGTTTGC CGATTAAAGA GGCAGAAGGC TTTGTATCC
    251  GTTTGAATGG AATCGCGCGC GGGGCTTTAG ACAGTAAATT CATGTTGAAG
    301  GCGGTAGCCA TAGATAAAGA TAAAAATCCT TTTATTATTA AGATGAATGA
    351  AAATCTAGTA ACCTTTATTT GCAAGAAGTC CGCCAGTTCG TGTAGTGACG
   401  GGCTGGATTA TTTTAAAGGA AATGATAAGG ACTGCAAGTT ACTTAAGTAG
  
```

This corresponds to the amino acid sequence <SEQ ID 822; ORF127-1>:

```

      1  MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN AVRAALLENNA
    51  HFMEKFYLN GRFKQTSTKW PSLPIKEAEG FCIRLNGIAR GALDSKFMLK
   101  AVAIDKDKNP FIIKMENLV TFICKKSASS CSDGLDYFKG NDKDCKLLK*
  
```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF127 shows 98.0% identity over a 150aa overlap with an ORF (ORF127a) from strain A of *N. meningitidis*:

```

      10      20      30      40      50      60
20  orf127.pep  MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN AVRAALLENNA HFMEKFYLN
      |||
20  orf127a     MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN TVRAALLENNA HFMEKFYLN
      |||
      10      20      30      40      50      60
25  orf127.pep  GRFKQTSTKW PSLPIKEAEG FCIRLNGIAR VARXALDSKFMLK AVAIDKDKNP FIIKMENLV
      |||
25  orf127a     GRFKQTSTKW PSLPIKEAEG FCIRLNGI-ARGALDSKFMLK AVAIDKDKNP FIIKMENLV
      |||
      70      80      90      100     110     120
30  orf127.pep  VTICKKSASS CSDGLDYFKG NDKDCKLLKX
      |||
30  orf127a     VTICKKSASS CSDGLDYFKG NDKDCKLLKX
      |||
      130     140     150
35  orf127.pep  VTICKKSASS CSDGLDYFKG NDKDCKLLKX
      |||
35  orf127a     VTICKKSASS CSDGLDYFKG NDKDCKLLKX
      |||
      120     130     140     150
  
```

The complete length ORF127a nucleotide sequence <SEQ ID 823> is:

```

      1  ATGACTGATA ATCGGGGGTT TACGCTGGTT GAATTAATAT CAGTGGTCTT
    51  GATATTGTCT GTACTTGCTT TAATTGTTTA TCCGAGCTAT CGCAATTATG
   101  TTGAGAAAGC AAAGATAAAT ACAGTGCGGG CAGCCTTGTT AGAAAATGCA
    151  CATTTTATGG AAAAGTTTTA TCTGCAGAAAT GGGAGATTTA AACAAACATC
    201  TACCAATGAG CCAAGTTTGC CGATTAAAGA GGCAGAAGGC TTTGTATCC
    251  GTTTGAATGG AATCGCGCGC GGGGCTTTAG ACAGTAAATT CATGTTGAAG
    301  GCGGTAGCCA TAGATAAAGA TAAAAATCCT TTTATTATTA AGATGAATGA
    351  AAATCTAGTA ACCTTTATTT GCAAGAAGTC CGCCAGTTCG TGTAGTGACG
   401  GGCTGGATTA TTTTAAAGGA AATGATAAGG ACTGCAAGTT ACTTAAGTAG
  
```

This encodes a protein having amino acid sequence <SEQ ID 824>:

```

      1  MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN TVRAALLENNA
    51  HFMEKFYLN GRFKQTSTKW PSLPIKEAEG FCIRLNGIAR GALDSKFMLK
   101  AVAIDKDKNP FIIKMENLV TFICKKSASS CSDGLDYFKG NDKDCKLLK*
  
```

ORF127a and ORF127-1 show 99.3% identity in 149 aa overlap:

```

      10      20      30      40      50      60
orf127a.pep  MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN TVRAALLENNA HFMEKFYLN
      |||
orf127-1     MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN AVRAALLENNA HFMEKFYLN
      |||
  
```

-473-

		10	20	30	40	50	60
		70	80	90	100	110	120
5	orf127a.pep	GRFKQTSTKWPSLPIKEAEGFCIRLNGIARGALDSKFMLKAVAIDKDKNPFIKMNENLV					
	orf127-1	GRFKQTSTKWPSLPIKEAEGFCIRLNGIARGALDSKFMLKAVAIDKDKNPFIKMNENLV					
10	orf127a.pep	130	140	150			
	orf127-1	130	140	150			

15 Homology with a predicted ORF from *N.gonorrhoeae*

ORF127 shows 97.3% identity over a 150 aa overlap with a predicted ORF (ORF127ng) from *N.gonorrhoeae*:

	orf127.pep	MTDNRGFTLVELISVVLILSVLALIVYPSYRNYVEKAKINAVRAALLENNAHFMEKFYLQN	60
20	orf127ng	MTDNRGFTLVELISVVLILSVLALIVYPSYRNYVEKAKINAVRAAFLENAHFMEKFYLQN	60
	orf127.pep	GRFKQTSTKWPSLPIKEAEGFCIRLNGIARXALDSKFMLKAVAIDKDKNPFIKMNENL	120
25	orf127ng	GRFKQTSTKWPSLPIKEAEGFCIRLNGI-ARGALDSKFMLKAVAIDKDKNPFIKMNENL	119
	orf127.pep	VTFICKKSASSCSDDLDFYFGNDKDKCKLLK	150
	orf127ng	VTFICKKSASSCSDDLDFYFGNDKDKCKLLK	149

The complete length ORF127ng nucleotide sequence <SEQ ID 825> is:

30	1	ATGACTGATA	ATCGGGGGTT	TACACTGGTT	GAATTAATAT	CAGTGGTCTT
	51	GATATTGTCT	GTAATTGCTT	TAATTGTTTA	TCCGAGCTAT	CGCAATTATG
	101	TTGAGAAAGC	AAAGATAAAT	GCAGTGCGGG	CAGCCTTGTT	AGAAAATGCA
	151	CATTTTATGG	AAAAGTTTTA	TCTGCAGAAT	GGGAGATTTA	AACAAACATC
35	201	TACCAAATGG	CCAAGTTTGC	CGATTAAAGA	GGCAGAAGGC	TTTGTATCC
	251	GTTTGAATGG	AATCGCGCGC	GGGGCTTTAG	ACAGTAAATT	CATGTTGAAG
	301	GCGGTAGCCA	TAGATAAAGA	TAAAATCCT	TTTATATTA	AGATGAATGA
	351	AAATCTAGTA	ACCTTTATTT	GCAAGAAGTC	CGCCAGTTTC	TGTAGTGACG
	401	GGCTGGATTA	TTTTAAAGGA	AATGATAAGG	ACTGCAAGTT	ACTTAAGTAG

This encodes a protein having amino acid sequence <SEQ ID 826>:

40	1	MTDNRGFTLV	ELISVVLILS	VLALIVYPSY	RNYVEKAKIN	AVRAAFLENA
	51	HFMEKFYLQN	GRFKQTSTKW	PSLPIKEAEG	FCIRLNGIAR	GALDSKFMLK
	101	AVAIDKDKNP	FIIKMNENLV	TFICKKSASS	CSDRLDYFKG	NDKDKCKLLK*

ORF127ng and ORF127-1 show 100.0% identity in 149 aa overlap:

		10	20	30	40	50	60
45	orf127-1.pep	MTDNRGFTLVELISVVLILSVLALIVYPSYRNYVEKAKINAVRAALLENNAHFMEKFYLQN					
	orf127ng-1	MTDNRGFTLVELISVVLILSVLALIVYPSYRNYVEKAKINAVRAALLENNAHFMEKFYLQN					
50	orf127-1.pep	70	80	90	100	110	120
	orf127ng-1	70	80	90	100	110	120
55	orf127-1.pep	130	140	150			
	orf127ng-1	130	140	150			

This analysis, including the fact that the predicted transmembrane domain is shared by the meningococcal and gonococcal proteins, suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

## 5 Example 98

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 827>

```
1  .GTGTCGCTGG CTTCCGGTGAT TGCCTCTCAA ATCTTCCTTT ACGAAGATTT
51  CAACCAAATG CCGAAAACCC GTGGAGCTAT CTGCGGTTT CTTGTCCAAT
101 ATTTATCTGG GGTTCAGCA GGGGTATTT GATTGAGTG CCGACGAGAA
151 CCCGTACTG CATATCTGGT CTTTGGCAGT AGAGGAACAG TATTACCTCC
201 TGTATCCCCT TTTGCTGATA TTTTGTGCA AAAAAACCAA ATCGCTACGG
251 GTGCTGCGTA ACATCAGCAT CATCCTGTTT TTGATTTTGA CTGCCTCATC
301 GTTTTGGCCA AGCGGGTTT ATACCGACAT CCTCAACCAA CCCAATACTT
351 ATTACCTTTC GACACTGAGG TTTCCCGAGC TGTGGCAGG TTCGCTGCTG
15 401 GCGGTTTACG GGCAAACGCA AAACGGCAGA CGGCAAACAG CAAATGGAAA
451 ACGGCAGTTG CTTTCATCAC TCTGCTTCGG CGCATTGCTT GCCTGCCTGT
501 TCGTGATTGA CAAACACAAT CCGTTTATCC CGGGAATGAC CCTGCTCCTT
551 CCCTGCCTGC TGACGGCACT GCTTATCCGG AGTATGCAAT ACGGGACACT
601 TCCGACCCGC ATCCTGTCGG CAAGCCCAT CGTATTTGTC GGCAAAATCT
20 651 CTTATTCCTT ATACCTGTAC CATTGGATTT TTATTGCTTT CGCTCCGCTC
701 ATTAGAGGCG GGAACAGCT CGGACTGCCT GCCG..
```

This corresponds to the amino acid sequence <SEQ ID 828; ORF128>:

```
1  .VSLASVIASQ IFLYEDFNQM RKTVELSAVF LSNIYLGQQ GYFDLSADEN
51  PVLHIWSLAV EEQYLLYPL LLIFCCCKTK SLRVLRNISI ILFLILTASS
25 101 FLPSGFYTDI LNQPNTYYLS TLRFPPELLAG SLLAVYGQTQ NGRRQTANGK
151 RQLSSLCFG ALLACLFVID KHNPFIPGMT LLLPCLLTAL LIRSMQYGTL
201 PTRILSASPI VFVGKISYSL YLYHWIFIAF APLIRGGKQL GLPA..
```

Further work revealed the complete nucleotide sequence <SEQ ID 829>:

```
1  ATGCAAGCTG TCCGATACAG ACCGGAAATT GACGGATTGC GGGCCGTCGC
30 51  CGTGCTATCC GTCATGATTT TCCACCTGAA TAACCGCTGG CTGCCCCGAG
101 GATTCCTGGG GGTGGACATT TTCTTTGTCA TCTCAGGATT CCTCATTACC
151 GGCATCATTC TTTCTGAAAT ACAGAACGGT TCTTTTCTT TCCGGGATTT
201 TTATACCCGC AGGATTAAGC GGATTTATCC TGCTTTTATT GCGGCCGTGT
251 CGCTGGCTTC GGTGATTGCC TCTCAAATCT TCCTTTACGA AGATTTCAAC
35 301 CAAATGCGGA AAACCGTGA GCTTCTGCG GTTTCTTGT CCAATATTTA
351 TCTGGGGTTT CAGCAGGGGT ATTTGATTT GAGTGCCGAC GAGAACCCCG
401 TACTGCATAT CTGGTCTTTG GCAGTAGAGG AACAGTATTA CCTCCTGTAT
451 CCCCTTTTGC TGATATTTTG CTGCAAAAAA ACCAAATCGC TACGGGTGCT
501 GCGTAACATC AGCATCATCC TGTTTTTGAT TTTGACTGCC TCATCGTTTT
40 551 TGCCAAGCGG GTTTTATACC GACATCCTCA ACCAACCCTA TACTTATTAC
601 CTTTCGACAC TGAGGTTTCC CGAGCTGTTG GCAGGTTTCG TGCTGGCGGT
651 TTACGGGCAA ACGCAAAACG GCAGACGGCA AACAGCAAAT GGAAAACGGC
701 AGTTGCTTTC ATCACTCTGC TTCGGCGCAT TGCTTGCCTG CCTGTTCGTG
751 ATTGACAAAC ACAATCCGTT TATCCCGGGA ATGACCCTGC TCCTTCCCTG
45 801 CTTGCTGACG GCACTCTTGA TCCGGAGTAT GCAATACGGG ACACTTCCGA
851 CCCGCATCCT GTCGGCAAGC CCCATCGTAT TTGTCGGCAA AATCTCTTAT
901 TCCCTATACC TGTACCATTT GATTTTTATT GCTTTCGCCC ATTACATTAC
951 AGGCGACAAA CAGCTCGGAC TGCTTGCCGT ATCGGCGGTT GCCGCTTGA
50 1001 CGGCGGATT TTCCCTGTTG AGTTATTATT TGATTGAACA GCCGCTTAGA
1051 AAACGGAAGA TGACCTTCAA AAAGGCATT TTCTGCCTCT ATCTCGCCCC
1101 GTCCCTGATA CTTGTCGGTT ACAACCTGTA CGCAAGGGGG ATATTGAAAC
1151 AGGAACACCT CCGCCCGTTG CCGGCGCGC CCCTTGCTG GGAATAATCAT
1201 TTCCGGGAAA CCGTCTGAC CCTCGGCGAC TCGCACGCCG GACACCTGAG
1251 GGGGTTTCTG GATTATGTCG GCAGCCGGGA AGGTGGGAAA GCCAAATACC
55 1301 TGTCCTCGA TTCGGAGTGT TTGGTTTGGG TAGATGAGAA GCTGGCAGAC
1351 AACCCTTAT GTCGAAAATA CCGGATGAA GTTGAAAAG CCGAAGCCGT
```



	1401	TTTCATTGCC	CAATTCTATG	ATTTGAGGAT	GGGCGGCCAG	CCTGTGCCGA
	1451	GATTGAAGC	GCAATCCTTC	CTAATACCCG	GGTCCCAGC	CCGATTCAGG
	1501	GAAACCGTCA	AAAGGATAGC	CGCCGTCAAA	CCCGTCTATG	TTTTTGCAAA
5	1551	CAACACATCA	ATCAGCCGTT	CGCCCCGTAG	GGAGGAAAAA	TTGAAAAGAT
	1601	TTGCCGCAAA	CCAATATCTC	CGCCCCATTG	AGGCTATGGG	CGACATCGGC
	1651	AAGAGCAATC	AGGCGGTCTT	TGATTTGATT	AAAGATATTC	CCAATGTGCA
	1701	TTGGGTGGAC	GCACAAAAAT	ACCTGCCCAA	AAACACGGTC	GAAATATACG
	1751	GCCGCTATCT	TTACGGCGAC	CAAGACCACC	TGACCTATTT	CGGTTCTTAT
10	1801	TATATGGGGC	GGGAATTCCA	CAAACACGAA	CGCCTGCTTA	AATCTTCCCA
	1851	CGGCGGCGCA	TTGCAGTAG			

This corresponds to the amino acid sequence <SEQ ID 830; ORF128-1>:

	1	<u>MQAVRYRPEI</u>	<u>DGLRAVAVLS</u>	<u>VMIFHLNNRW</u>	<u>LPGGFLGVDI</u>	<u>FFVISGFLIT</u>
	51	<u>GIILSEIQNG</u>	<u>SFSFRDFYTR</u>	<u>RIKRIYPAFI</u>	<u>AAVSLASVIA</u>	<u>SOIFLYEDFN</u>
15	101	<u>QMRKTVELSA</u>	<u>VFLSNIIYLG</u>	<u>QQGYFDLSAD</u>	<u>ENPVLHWSL</u>	<u>AVEEQYLLY</u>
	151	<u>PLLLIFCCKK</u>	<u>TKSLRVLRLN</u>	<u>SIILFLILTA</u>	<u>SEFLSGHYFT</u>	<u>DILNQPNTRY</u>
	201	<u>LSTLRFPPELL</u>	<u>AGSLLAVYQG</u>	<u>TQNGRRQTAN</u>	<u>GKROLSSSLC</u>	<u>FGALLACLFV</u>
	251	<u>IDKHNPFIPT</u>	<u>MTLLLPCLLT</u>	<u>ALLIRSMQYG</u>	<u>TLPTRILSAS</u>	<u>PIVFGVKISY</u>
	301	<u>SLYLYHWIFI</u>	<u>AFAHYITGDK</u>	<u>QLGLPAVSAV</u>	<u>AALTAFGSLL</u>	<u>SYYLIEQPLR</u>
20	351	<u>KRKMTFKKAF</u>	<u>FCLYLAPSLI</u>	<u>LVGYNLYARG</u>	<u>ILKQEHLRPL</u>	<u>PGWLAENHN</u>
	401	<u>FPETVTLTGD</u>	<u>SHAGHLRGFL</u>	<u>DYVGSREGWK</u>	<u>AKILSLDSEC</u>	<u>LGVVDEKLAD</u>
	451	<u>NPLCRKYRDE</u>	<u>VEKAEAVFIA</u>	<u>QFYDLRMGGQ</u>	<u>PVPRFEAQSF</u>	<u>LIPGFPARFR</u>
	501	<u>ETVKRIAFAV</u>	<u>PVYVFANNTS</u>	<u>ISRSPLREEK</u>	<u>LKRFAANQYL</u>	<u>RPIQAMGDIG</u>
	551	<u>KSNQAVFDLI</u>	<u>KDIPNVHWVD</u>	<u>AQKYLPKNTV</u>	<u>EIYGRYLYGD</u>	<u>QDHLTYFGSY</u>
	601	<u>YMGREFHKHE</u>	<u>RLKSSHGGA</u>	<u>LQ*</u>		

25 Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical integral membrane protein HI0392 of *H.influenzae* (accession number U32723)

ORF128 and HI0392 show 52% aa identity in 180aa overlap:

	Orf128: 1	VSLASVIASQIFLYEDFNQMRKTVELSAVFLSNIYLGFGQQGYFDLSADENPVLHIWSLAV	60
30		++L S IAS IF+Y DFN++RKT+EL+ FLSN YLG QGYFDLSA+ENPVLHIWSLAV	
	HI0392: 46	MALVSFIASAIIFIYDNDFNKLKRTIELAIAFLSNFYLGTLTQGYFDLSANENPVLHIWSLAV	105
	Orf128: 61	EEQXXXXXXXXXIFCCKKTKSLRVLNRNISIILFLILTASSFLPSGFYTDILNPQNTYYLS	120
35		E Q I KK + ++VL I++ILF IL A+SF+ + FY ++L+QPN YYLS	
	HI0392: 106	EGQYYLIFFLILILAYKKKFREVVKVLFITLILFFILLATSFVSANFYKEVLHQPNYYLS	165
	Orf128: 121	TLRFPELLAGSLLAVYGTQNGRRQTANGKRQLLSSLCFGALLACLFLVIDKHNPFPFGMT	180
		LRFPELL GSLLA+Y N + Q + +L+ L L +CLF+++ + FIPG+T	
	HI0392: 166	NLRFPELLVGSLLAIYHNLSN-KVQLSKQVNNILAILSTLLFLSCFLFMNNNIAFIPGIT	224

40 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF128 shows 98.0% identity over a 244aa overlap with an ORF (ORF128a) from strain A of *N.*

*meningitidis:*

[illegible]

-476-

5 orf128.pep RQLSSSLCFGALLACLFVIDKHNPFIPGMTLLLPCLLTALLIRSMQYGTLPTRILSASPI  
 orf128a RQLSSSLCFGALLACLFVIDKHNPFIPGMTLLLPCLLTALLIRSMQYGTLPTRILSASPI  
 10 orf128.pep VFVVGKISYSLSLYLHWIFIAFAPLIRGGKQLGLPA  
 orf128a VFVVGKISYSLSLYLHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYLIEQPLRKR  
 orf128a KMTFFKKAFFCLYLAPSLILVGYNLYARGILKQEHRLPLPGAPLAAENHFPETVLTGLGDSH

15 The complete length ORF128a nucleotide sequence <SEQ ID 831> is:

1 ATGCAAGCTG TCCGATACAG ACCGGAAATT GACGGATTGC GGGCCGTCGC  
 51 CGTGCTATCC GTCATGATTT TCCACCTGAA TAACCGCTGG CTGCCCGGAG  
 101 GATTCTTGGG GGTGGACATT TTCTTTGTCA TCTCAGGATT CCTCATTACC  
 151 GGCATCATTC TTTCTGAAAT ACAGAACGGT TCTTTTCTT TCCGGGATTT  
 201 TTATACCCGC AGGATTAAGC GGATTTATCC TGCTTTTAT GCGGCCGTGT  
 251 CCTGGGCTTC GGTGATTGCC TCTCAAATCT TCCTTTACGA AGATTTCAAC  
 301 CAAATGCGGA AAACCGTGGA GCTTCTGCG GTTTCTTGT CCAATATTTA  
 351 TCTGGGGTTT CAGCAGGGGT ATTTTCGATT GAGTGCCGAC GAGAACCCCG  
 401 TACTGCATAT CTGGTCTTTG GCAGTAGAGG AACAGTATTA CCTCCTGTAT  
 451 CCTCTTTTGC TGATATTTTG CTGCAAAAAA ACAAATCGC TACGGGTGCT  
 501 GCGTAACATC AGCATCATCC TATTTCTGAT TTTGACTGCC ACATCGTTTT  
 551 TGCCAAGCGG GTTTTATACC GATATTCTCA ACCAACCCTA TACTTATTAC  
 601 CTTTCGACAC TGAGGTTTCC CGAGCTGTTG GCAGGTTCCG TGCTGGCGGT  
 651 TTACGGGCAA ACGCAAAACG GCAGACGGCA AACAGCAAAT GGAAAACGGC  
 701 AGTTGCTTTC ATCATTCTGC TTCGGCGCAT TGCTGCTCTG CCTGTTCTGT  
 751 ATTGACAAAC ACAATCCGTT TATCCCGGGA ATGACCCTGC TCCTTCCTGT  
 801 CCGCTGACG GCACCTGCTTA TCCGGAGTAT GCAATACGGG ACACTTCOGA  
 851 CCCGCATCCT GTCGGCAAGC CCCATCGTAT TTGTCGGCAA AATCTCTTAT  
 901 TCCCTATACC TGTACCATTG GATTTTTTAT GCTTTCGCCC ATTACATTAC  
 951 AGGCGACAAA CAGCTCGGAC TGCTGCGGT ATCGGCGGT GCGCGTTGA  
 1001 CGGCCGGATT TTCCCTGTTG AGTTATTATT TGATTGAACA GCCGCTAGA  
 1051 AAACGGAAGA TGACCTTCAA AAAGGCATTT TTCTGCCTCT ATCTCGCCCC  
 1101 GTCCCTGATA CTTGTGCGGT ACAACCTGTA CGCAAGGGGG ATATTGAAC  
 1151 AGGAACACCT CCGCCCGTTG CCCGGCGCGC CCCTTGCTGC GGAATATCAT  
 1201 TTTCCGAAA CCGTCTGAC CCGCGCGGAC TCGCACGCGG GACACCTGCG  
 1251 GGGGTTTCTG GATTATGTCG GCAGCCGGGA AGGGTGGAAG GCCAAATCC  
 1301 TTCCCTCGA TTCGGAGTGT TTGGTTTGGG TAGATGAGAA GCTGGCAGAC  
 1351 AACCCGTTAT GTCGAAAATA CCGGATGAA GTTGAAAAG CCGAAGCCGT  
 1401 TTTCATTGCC CAATTCTATG ATTTGAGGAT GGGCGGCCAG CCCGTGCCGA  
 1451 GATTGGAAGC GCAATCTTTC CTAATACCG GGTCCCAGC CCGATTGAGG  
 1501 GAAACCGTCA AAAGGATAGC CGCCGTCAA CCCGTCTATG TTTTGCAAA  
 1551 CAACACATCA ATCAGCCGTT CGCCCTGAG GGAGGAAAAA TTGAAAAGAT  
 1601 TTGCCGAAA CCAATATCTC CGCCCCATTC AGGCTATGGG CGACATCGGC  
 1651 AAGAGCAATC AGGCGTCTT TGATTTGATT AAAGATATTC CCAATGTGCA  
 1701 TTGGGTGGAC GCACAAAAAT ACCTGCCCAA AACACGGTC GAAATATACG  
 1751 CGCGCTATCT TTACGGCGAC CAAGACCACC TGACCTATTT CGGTTCTTAT  
 1801 TATATGGGGC GGGAAATTTCA CAAACACGAA CGCTGCTTA AATCTTCTCG  
 1851 CGACGGCGCA TTGCAGTAG

This encodes a protein having amino acid sequence <SEQ ID 832>:

55 1 MQAVRYRPEI DGLRAVAVLS VMIFHLNNRW LPGGFLGVDI FFVISGFLIT  
 51 GIILSEIQNG SFSFRDFYTR RIKRIYPAFI AAVSLASVIA SQIFLYEDFN  
 101 QMRKTVELSA VFLSNIYLG F QQGYFDLSAD ENPVLHIWSL AVEEQYLLY  
 151 PLLLIFCCKK TKSLRVLRLNI SIILFLILTA TSFLPSGFYT DILNQPNYY  
 201 LSTLRFPELL AGSL LAVYQ TQNGRRQTAN GKRQLSSSLC FGALLACLFV  
 60 251 IDKHNPFIPG MTL L L PCLLT ALLIRSMQYGTLPTRILSAS PIVFVGKISY  
 301 SLYLYHWIFT AFAHYITGDK QLGLPAVSAV AALTAGFSLL SYYLIEQPLR  
 351 KRKMTFFKAF FCLYLAPSLI LVGYNLYARG ILKQEHRLPL PGAPLAAENH  
 401 FPETVLTGLD SHAGHLRGFL DYVGSREGWK AKILSLDSEC LVWVDEKLAD  
 451 NPLCRKYRDE VEKAEAVFIA QFYDLRMGGQ PVPRFEAQSF LIPGFPARFR  
 65 501 ETVKRIAAVK PVYVFANNTS ISRSPLREEK LKRFANQYL RPIQAMGDIG  
 551 KSNQAVFDLI KDIPNVHWD AQKYLKNTV EIYGRYLYGD QDHLTYFGSY

601 YMGREFHKHE RLLKSSRDGA LQ\*

ORF128a and ORF128-1 show 99.5% identity in 622 aa overlap:

5 orf128a.pep MQAVRYRPEIDGLRAVAVLSVMIFHLNNRWLPGGFLGVDIFFVISGFLITGIILSEIQNG  
 orf128-1 MQAVRYRPEIDGLRAVAVLSVMIFHLNNRWLPGGFLGVDIFFVISGFLITGIILSEIQNG

10 orf128a.pep SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTVELSAVFLSNIYLG  
 orf128-1 SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTVELSAVFLSNIYLG

15 orf128a.pep QQGYFDLSADENPVLHIWLSLAVEEQYLLYPLLLIFCCKKTKSLRVLRNISIILFLILTA  
 orf128-1 QQGYFDLSADENPVLHIWLSLAVEEQYLLYPLLLIFCCKKTKSLRVLRNISIILFLILTA

20 orf128a.pep TSFLPSGFYTDILNQPNYYLSTLRFPPELLAGSLLAVYGQTQNGRRQTANGKRQLLSSLC  
 orf128-1 TSFLPSGFYTDILNQPNYYLSTLRFPPELLAGSLLAVYGQTQNGRRQTANGKRQLLSSLC

25 orf128a.pep FGALLACLFVIDKHNPFI PGMTLLLPCLLTALLIRSMQYGTLPTRILSASP VFGKISY  
 orf128-1 FGALLACLFVIDKHNPFI PGMTLLLPCLLTALLIRSMQYGTLPTRILSASP VFGKISY

30 orf128a.pep SLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYLLIEQPLRKRKMTFKKAF  
 orf128-1 SLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYLLIEQPLRKRKMTFKKAF

35 orf128a.pep FCLYLAPSLILVGYNLYARGILKQEHRLPLPGAPLAENHFPETVLTGLGDSHAGHLRGFL  
 orf128-1 FCLYLAPSLILVGYNLYARGILKQEHRLPLPGAPLAENHFPETVLTGLGDSHAGHLRGFL

40 orf128a.pep DYVGSREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ  
 orf128-1 DYVGSREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ

45 orf128a.pep PVPRFEAQSFILPGFPARFRET VKRIA AVKPVYVFANNTSISRSPREEKLRFAANQYL  
 orf128-1 PVPRFEAQSFILPGFPARFRET VKRIA AVKPVYVFANNTSISRSPREEKLRFAANQYL

orf128a.pep RPIQAMGDIGKSNQAVFDLIK DIPNVHWVDAQKYL PKNTVEIYGRYLYGDQDHLTYFGSY  
 orf128-1 RPIQAMGDIGKSNQAVFDLIK DIPNVHWVDAQKYL PKNTVEIYGRYLYGDQDHLTYFGSY

orf128a.pep YMGREFHKHERLLKSSRDGALQX  
 orf128-1 YMGREFHKHERLLKSSSHGGALQX

Homology with a predicted ORF from *N.gonorrhoeae*

ORF128 shows 93.4% identity over 244 aa overlap with a predicted ORF (ORF128ng) from *N. gonorrhoeae*:

50 orf128.pep VSLASVIASQIFLYEDFNQMRKTVELSAV 30  
 orf128ng ILSEIQNGSFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTIELSTVF 112

55 orf128.pep LSNIYLGFFQQGYFDLSADENPVLHIWLSLAVEEQYLLYPLLLIFCCKKTKSLRVLRNISI 90  
 orf128ng LSNIYLGFRLLGYFDLSADENPVLHIWLSLAVEEQYLLYPLLLIFCYKTKSLRVLRNISI 172

60 orf128.pep ILFLILTASSFLPSGFYTDILNQPNYYLSTLRFPPELLAGSLLAVYGQTQNGRRQTANGK 150  
 orf128ng ILFLILTASSFLPAGFYTDILNQPNYYLSTLRFPPELLVGSLLAVYGQTQNGRRQTENGK 232

orf128.pep RQLLSSLCFGALLACLFVIDKHNPFI PGMTLLLPCLLTALLIRSMQYGTLPTRILSASPI 210  
 orf128ng RQLLSSLCFGALLVCLFVIDKHDFIPGITLLLPCLLTALLIRSMQYGTLPTRILSASPI 292

orf128.pep VFVGKISYSLYLYHWIFIAFAPLIRGGKQLGLPA 244  
|||||  
orf128ng VFVGKISYSLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYLLIEQPLRKR 352

5 The complete length ORF128ng nucleotide sequence <SEQ ID 833> is:

1 ATGCAAGCTG TCCGATACAG GCCTGAAATT GACGGATTGC GGGCCGTCGG  
51 CGTGCTATCC GTCATTATTT TCCACCTGAA TAACCGCTGG CTGCCCGGAG  
101 GATTCTCTGGG GGTGGACATT TTCTTTGTCA TCTCGGGATT CCTCATTACC  
10 151 AACATCATT TTTCTGAAAT ACAGAACGGT TCTTTTCTT TCCGGGATTT  
201 TTATACCCGC AGGATTAAGC GGATTTATCC TGCTTTTATT GCGGCCGTGT  
251 CCCTGGCTTC GGTGATTGCT TCTCAAATCT TCCTTTACGA AGATTTCAAC  
301 CAAATGAGGA AAACCATAGA GCTTTCTACG GTTTTTTTGT CCAATATTTA  
351 TTTGGGGTTC CGATTGGGGT ATTTGATTG GAGTGCCGAC GAGAACCCCG  
401 TACTGCATAT CTGGTCTTTG GCGGTAGAGG AACAGTATTA CCTCCTGTAT  
15 451 CCTCTTTTGC TGATATCTGT TTACAAAAA ACCAAATCAC TACGGGTGCT  
501 GCGTAATATC AGCATCATCC TGTTTCTGAT TTTGACCGCA TCATCGTTTT  
551 TGCCGGCCCGG TTTTATACG GACATCCTCA ACCAACCcaa TACTTATTAC  
601 CTTTCGACAC TGAGGTTTCC CGAGCTGTTG GTGGGTTTCG TGTTGGCGGT  
651 TTACGGGCAA ACGCAAAACG GCAGACGGCA AACAGAAAAT GGAAACCGCG  
20 701 AGTTGCTTTC ATTACTCTGT TTCGGCGCat tgCTTGTCTG CCTGTTCTGT  
751 ATGACAAAAC CAGATCCGTT TATCCCGGGA ATAACCCGTC TCCTTCCCTG  
801 CCTGCTGACG GCGCTGCTTA TCCGGAGTAT GCAATACGGG AACTTCCGA  
851 CCCGCATCCT GTCGGCAAGC CCCATCGTAT TTGTCGGCAA AATCTCTTAT  
901 TCCCTATACC TGTACATTG GATTTTATT GCCTTCGCC ATTACATTAC  
25 951 AGCGACAAA CAGCTCGGAC TGCTGCGGT ATCGCGGTT GCCGCGTTGA  
1001 CGGCCGGATT TTCCTGTTG AGCTATTATT TGATGAACA GCCGCTTAGA  
1051 AAACGGAAGA TGACCTTCAA AAAGGCATT TTCTGCCTTT ATCTCGCCCC  
1101 TGCCCTGATG CAGGAGTAGT ACAACCTGTA TTCAAGAGGG ATATTGAAC  
1151 AGGAACACCT CCGCCCGCTG CCCGGCACGC CCGTTGCTGC GGAAATAAT  
30 1201 TTTCCGGAAG CCGTCTTGAC CCTCGCGCAC TCGCACGCCG GACACCTGCG  
1251 GGGGTTTCTG GATTATGTCG GCGGCAGGGA AGGGTGGAAG GCTAAATCC  
1301 TGTCCCTCGA TTCGGAGTGT TTGGTTTGGG TGGATGAGAA GCTGGCAGAC  
1351 AACCCTGTTG GCCGAAAATA CCGGGATGAA GTTGAAAAG CCGAAGCTGT  
1401 TTTCAATTGCC CAATCTATG ATTTGAGGAT GGGCGGCCAG CCCGTGCCGA  
35 1451 GATTTGAAGC GCAATCCTTC CTGATACCCG GGTCAAAGC CCGATTCAAG  
1501 GAAACCGTCA AGAGGATAGC CGCCGTCAA CCTGTATATG TTTTGGAAA  
1551 CAATACATCA ATCAGCCGTT CTCCCTTGAG GGAGGAAAAA TTGAAAAGAT  
1601 TTGCTATAAA CCAATACCTC CGGCCTATTC GGGCTATGGG CGACATCGGC  
1651 AAGAGCAATC AGGCGGTCTT TGATTGGTT AAAGATATTC CCAATGTGCA  
40 1701 TTGGGTGGAC GCACAAAAAT ACCTGCCCAA AAACACGGTC GAAATACACG  
1751 GACGCTATCT TTACGGCGAC CAAGACCACC TGACCTATTT CGGTTCTTAT  
1801 TATATGGGGC GGAATTTCA CAAACACGAA CGCTGCTCA AGCATCCCCG  
1851 AGGCGGCGCA TTGCAGTAG

This encodes a protein having amino acid sequence <SEQ ID 834>:

45 1 MQAVRYRPEI DGLRAVAVLS VIIFHLNNRW LPGGFLGVDI FFVISGFLIT  
51 NIILSEIQNG SFSFRDFYTR RIKRIYPAFI AAVSLASVIA SQIFLYEDFN  
101 QMRKTIELST VFLSNIYLG FRLGYFDLSAD ENPVLHIWSL AVEEQYLLY  
151 PLLLIFCYKK TKSLRVLRNI SIILFLILTA SSFLPAGFYT DILNQPNTRY  
50 201 LSTLRFPELL VGSLLAVYGQ TQNGRRQ TEN GKRQLLSLLC FGALLVCLFV  
251 IDKHDPFPIG ITLLLPCLLT ALLIRSMQYG TLPTRILSAS PIVFVGKISY  
301 SLYLYHWIFI AFAHYITGDK QLGLPAVSAV AALTAGFSL SYLLIEQPLR  
351 KRKMTFKKAF FCLYLAPSLM LVGYNLYSRG ILKQEHRLPL PGTPVAAENN  
401 FPETVLTLGD SHAGHLRGFL DYVGGREGWK AKILSLDSEC LVWVDEKLAD  
55 451 NPLCRKYRDE VEKAEAVFIA QFYDLRMGGQ PVPRFEAQSF LIPGFKARFR  
501 ETVKRIAIVK PVYVFANNTS IRSPLREEK LKRFAINQYL RPIRAMGDIG  
551 KSNQAVFDLV KDIPNVHWDV AQKYLPKNTV EIHGRYLYGD QDHLTYFGSY  
601 YMGREFHKHE RLLKHSRGG A LQ\*

ORF128ng and ORF128-1 show 95.7% identity in 622 aa overlap:

60 orf128-1.pep MQAVRYRPEIDGLRAVAVLSVMIFHLNNRWLPGGFLGVDIFFVISGFLITGIILSEIQNG  
|||||  
orf128ng MQAVRYRPEIDGLRAVAVLSVIIFHLNNRWLPGGFLGVDIFFVISGFLITNIILSEIQNG  
orf128-1.pep SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTVELSAVFLSNIYLG

5 orf128ng |||||SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTIELSTVFLSNIYLG  
orf128-1.pep QQGYFDLSADENPVLHIWSLAVEEQYLLYPLLIFCCKKTKSLRVLNRNISIILFLILTA  
6 orf128ng : |||||RLGYFDLSADENPVLHIWSLAVEEQYLLYPLLIFCYKTKSLRVLNRNISIILFLILTA  
10 orf128-1.pep SSFLPSGFYTDILNQPNNTYYLSTLRFPPELLAGSLLAVYGQTQNGRRQTANGKRQLSSLC  
orf128ng |||||SSFLPAGFYTDILNQPNNTYYLSTLRFPPELLVGSLLAVYGQTQNGRRQTENGKRQLSSLLC  
15 orf128-1.pep FGALLACLFVIDKHNPFI PGMTLLLPCLLTALLIRSMQYGTLPTRILSASP I VFVGKISY  
orf128ng |||||FGALLVCLFVIDKHDPFI PGITLLLPCLLTALLIRSMQYGTLPTRILSASP I VFVGKISY  
20 orf128-1.pep SLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYYLIEQPLRKRKMTFKKAF  
orf128ng |||||SLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYYLIEQPLRKRKMTFKKAF  
25 orf128-1.pep FCLYLAPSLILVGYNLYARGILKQEHRLPLPGAPLAAENHFPETVLTGLGDSHAGHLRGFL  
orf128ng |||||FCLYLAPSLMLVGYNLYSRGILKQEHRLPLPGTPVAAENHFPETVLTGLGDSHAGHLRGFL  
30 orf128-1.pep DYVGSREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ  
orf128ng |||||DYVGGREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ  
35 orf128-1.pep PVPREFEASFLIPGFPARFRET VKRIA AVKPVYVFANNTSISRSPLREEKLKRFAANQYL  
orf128ng |||||PVPREFEASFLIPGFKARFRET VKRIA AVKPVYVFANNTSISRSPLREEKLKRFAINQYL  
40 orf128-1.pep RPIQAMGDIGKSNQAVFDLIK DIPNVHWVDAQKYL PKNTVEIYGRYLYGDQDHLTYFGSY  
orf128ng |||||RPIRAMGDIGKSNQAVFDLVKDIPNVHWVDAQKYL PKNTVEIHGRYLYGDQDHLTYFGSY  
45 orf128-1.pep YMGREFHKHERLLKSSHGGALQX  
orf128ng |||||YMGREFHKHERLLKHSRGGALQX  
610 620

40 In addition, ORF218ng shows homology to a hypothetical *H.influenzae* protein:

sp|P43993|Y392\_HAEIN HYPOTHETICAL PROTEIN HI0392 >gi|1074385|pir||B64007  
hypothetical protein HI0392 - Haemophilus influenzae (strain Rd KW20)  
50 >gi|1573364 (U32723) H. influenzae predicted coding region HI0392 [Haemophilus  
influenzae] Length = 245  
Score = 239 bits (604), Expect = 3e-62  
Identities = 124/225 (55%), Positives = 152/225 (67%), Gaps = 1/225 (0%)  
Query: 38 VDIFFVISGFLITNIILSEIQNGSFSFRDFYTRRIKRIYPXXXXXXXXXXXXXXFLYE 97  
+DIFFVISGFLIT II++EIQ SFS + FYTRRIKRIYP F+Y  
Sbjct: 1 MDIFFVISGFLITGIIITEIQNSFSLKQFYTRRIKRIYPAFITVMALVSFIASAIFIYN 60  
Query: 98 DFNQMRKTIELSTVFLSNIYLGFR LGYFDLSADENPVLHIWSLAVEEQXXXXXXXXXIFC 157  
DFN++RKTIEL+ FLSN YLG GYFDLSA+ENPVLHIWSLAVE Q I  
Sbjct: 61 DFNKLRKTIELAI AFLSNFYLG LTQGYFDLSANENPVLHIWSLAVEGQYLLIFPLILILA 120  
55 Query: 158 YKTKSLRVLNRNISIILFLILTASSFLPAGFYTDILNQPNNTYYLSTLRFPPELLVGSLLAV 217  
YKK + ++VL I++ILF IL A+SF+ A FY ++L+QPN YYLS LRFPELLVGSLLA+  
Sbjct: 121 YKKFREVKVLFIIITLILFFILLATS FVSANFYKEVLHQPNIIYLSNLRFPELLVGSLLAI 180  
60 Query: 218 YGQTQNGRRQTENGKRQLSLLCFGALLVCLFVIDKHDPFI PGIT 262  
Y N + Q +L++L L CLF+++ + FIPGIT  
Sbjct: 181 YHNLSN-KVQLSKQVNNILAILSTLLLFSCFLMNNNIAFIPGIT 224

This analysis, including the identification of several putative transmembrane domains, suggests that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 99

- 5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 835>:

```

1  ..ATTATTACG AATACCGCTG GATGTTTCTT TACGGCGCAC TGACGACCTT
51  GGGGCTGACG GTCGTGGCAA C.GCGGGCGG TTCGGTATTG GGTCTGTTGT
101 TGGCGTTGGC GCGCCTGATT CACTTGGAAA AAGCCGGTGC GCCGATGCCG
151 GTGCTGGCGT GGGCGTTGCC TAAAGTTTCG CTGCTGTATG TTACGCTGTT
10  201 CCGGGGTACG CCGCTGTTTG TGCAGATTGT GATTTGGGCG TATGTGTGGT
251 TTCCGTTTTT CGTC..

```

This corresponds to the amino acid sequence <SEQ ID 836; ORF129>:

```

1  ..IIYEYRWMFL YGALTTLGLT VVAXAGGSVL GLLLALARLI HLEKAGAPMR
51  VLAWALRKVS LLYVTLFRGT PLFVQIVIWA YVWFPPFFV..

```

- 15 Further work revealed the complete nucleotide sequence <SEQ ID 837>:

```

1  ATGGATTTTC GTTTTCACAT TATTTACGAA TACCGCTGGA TGTTCCTTTA
51  CCGCGCACTG ACGACCTTGG GGCTGACGGT CGTGGCAACG GCGGGCGGTT
101 CCGTATTGGG TCTGTTGTTG GCGTTGGCGC GCCTGATTCA CTTGGAAAAA
151 GCCGGTGCGC CGATGCGCGT GCTGGCGTGG GCGTTGCGTA AAGTTTCGCT
20  201 GCTGTATGTT ACGCTGTTCC GGGGTACGCC GCTGTTTGTG CAGATGTGTA
251 TTTGGGCGTA TGTGTGGTTT CCGTTTTTCG TCCATCCTTC AGACGGCATT
301 TTGGTCAGCG GCGAGGCGGC AATCGCGCTG CGTCGCGGAT ACGGGCCGCT
351 GATTGCCGGT TCTTTGGCAC TGATCGCCAA CTCGGGGGCG TATATCTGTG
401 AGATTTTCCG CGCGGGCATC CAGTCTATAG ACAAAGGACA GATGGAGGCG
25  451 GCGCGTTCTT TGGGGCTGAC CTATCCGCAG GCGATGCGCT ATGTGATTCT
501 GCCGCAGGCA TTGCGCCGCA TGCTGCCGCC TTTGGCGAGC GAGTTCATCA
551 CGCTCTTGAA AGACAGCTCG CTGCTGTCGG TCATTGCTGT GCGCGAGTTG
601 GCGTATGTTT AGAATACGAT TACGGGCCGG TATTCGGTTT ATGAAGAACC
651 GCTTTACACC GTCGCCCTGA TTTATCTGTT GATGACGACT TTCTTAGGCT
30  701 GGATATTCTT CCGTTTGCAA AACGTTACA ATCCGCAACA CCGCTGA

```

This corresponds to the amino acid sequence <SEQ ID 838; ORF129-1>:

```

1  MDFRFDIIYE YRWMFLYGAL TTLGLTVVAT AGGSVLGLLL ALARLIHLEK
51  AGAPMRVLAW ALRKVSLLYV TFRGTPLFV QIVIWAYVWF PFFVHPSDGI
101 LVSGEAAIAL RRGYGPLIAG SLALIANSGA YICEIFRAGI QSIDKGQMEA
35  151 ARSLGLTYPQ AMRYVILPQA LRRMLPLAS EFITLLKDSS LLSVIAVAEL
201 AYVQNTITGR YSVYEEPLYT VALIYLLMTT FLGWIFLRLE KRYNPQHR*

```

Computer analysis of this amino acid sequence gave the following results:

### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF129 shows 98.9% identity over a 88aa overlap with an ORF (ORF129a) from strain A of *N.*

- 40 *meningitidis*:

```

                                     10    20    30    40    50
orf129.pep      IIYEYRWMFLYGALTTLGLTVVAXAGGSVLGLLLALARLIHLEKAGAPMRVLAW
                |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf129a      MDFRFDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW
                10    20    30    40    50    60

                                     60    70    80
orf129.pep      ALRKVSLLYVTLFRGTPLFVQIVIWAYVWFPPFFV
                |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

orf129a      ALRKVSLLYVTLFRGTFLFVQIWIWAYVWFFVFVHPSDGILVSGEAAIALRRGYGPLIAG

5 orf129a SLALIANSGAYICEIFRAGIQSIDKGQMEAARSGLTYPQAMRYVILPQALRRMLPPLAS  
130 140 150 160 170 180

The complete length ORF129a nucleotide sequence <SEQ ID 839> is:

	1	ATGGATTTC	GTTTTGACAT	TATTACGAA	TACCCTGGA	TGTTCTTTA
	51	CGGCGACTG	ACGACCTTG	GGCTGACGT	CGTGCGACG	CGGGGCGGT
10	101	CGGTATTGGG	TCTGTTGTT	GCGTTGGCGC	GCCTGATTCA	CTTGAAAAA
	151	GCCGCTGCG	CGATGCGGT	GCTGGCGTG	CGCTGCGTA	AGGTTTCCG
	201	GCTGTATGTT	ACGCTGTTCC	GGGGTACGCC	GCTGTTTGTTG	CAGATTGTGA
	251	TTTGGGCGTA	TGTGTGTTT	CCGTTTTTCG	TCCATCCTTC	AGACGGCATT
	301	TTGTTTAGCG	GCGAGGCGGC	AATCGCGCTG	CGTCGCGGAT	ACGGGCGCT
15	351	GATGCGCGT	TCTTTGGCAC	TGATCGCCAA	CTCGGGGCGC	TATATCTGTG
	401	AGATTTTCCG	CGCGGGCATC	CAGTCTATAG	ACAAAGGACA	GATGGAGCGC
	451	GCGCGTCTTT	TGGGGCTGAC	CTATCCGCAG	CGGATGCGCT	ATGTTGATTCT
	501	GCCGCAGGCA	TTGCGCGTA	TGCTGCCGCC	TTTGCGCAGC	GAGTTTCATCA
	551	CGCTCTTGAA	AGACAGCTCG	CTGCTGTCGG	TCATTGCTGT	GGCGGAGTTG
20	601	GCGTATGTTC	AGAATACCAT	TACGGGCCCG	TATTGCGTTT	ATGAAGAACC
	651	GCTTTACACC	GTCGCCCTGA	TTTATCTGTT	GATGACGACT	TTCTTAGGCT
	701	GGATATTCTCT	GCGTTTGAA	AAACGTTACA	ATCCGCAACA	CCGCTGA

This encodes a protein having amino acid sequence <SEQ ID 840>:

25

1	MDFRFDDIYE	YRWMFLYGAL	TTLGLTVVAT	AGGSVLGLLL	ALARLIHLEK
51	AGAPMRVLAW	ALRKVSLLYV	TLFRGTPLEV	QIVIWAYVWF	FFVHPSDGI
101	LVSGEAAIAL	RRGYGPLIAG	SLALIANSGA	YICIFRAGI	QSIDKQMEA
151	ARSLGLTPQ	AMRYVILPQA	LRRMLPLAS	EFITLLKDSS	LLSVIAVAEL
201	AYVONTITGR	YSVYEEPLYT	VALIYLLMTT	FLGWIFLRL	KRYNPQHR*

ORF129a and ORF129-1 show 100.0% identity in 248 aa overlap:

30	orf129a.pep	MDFRFDIIIEYRWMFLYGALTTTGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW 
	orf129-1	MDFRFDIIIEYRWMFLYGALTTTGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW 
35	orf129a.pep	ALRKVSLLYVTLFRGTPLFVQIVIWAYVWFPPFVHPSDGILVSGEAAIALRRGYGPLIAG 
	orf129-1	ALRKVSLLYVTLFRGTPLFVQIVIWAYVWFPPFVHPSDGILVSGEAAIALRRGYGPLIAG 
40	orf129a.pep	SLALIANS GAYICEIFRAGIQSIDKGQMEAARSLGLTYPQAMRYVILPQALRRMLPPLAS 
	orf129-1	SLALIANS GAYICEIFRAGIQSIDKGQMEAARSLGLTYPQAMRYVILPQALRRMLPPLAS 
45	orf129a.pep	EFITLLKDSSLLSVIAVAELAYVQNTITGRYSVYEEPLYTVALIYLLMTTFLGWIFLRLE 
	orf129-1	EFITLLKDSSLLSVIAVAELAYVQNTITGRYSVYEEPLYTVALIYLLMTTFLGWIFLRLE 

Homology with a predicted ORF from *N.gonorrhoeae*

50 ORF129 shows 98.9% identity over a 88 aa overlap with a predicted ORF (ORF129ng) from  
*N.gonorrhoeae*:

	orf129.pep	IIEYRWMFLYGALTTLGLTVVAXAGGSVLGLLALARLIHLEKAGAPMRVLAW	54
55	orf129ng	MDFRFDIIEYRWMFLYGALTTLGLTVVATAGGSVLGLLALARLIHLEKAGAPMRVLAW	60
	orf129.pep	ALRKVSLLYVTLFRGTPLFVQIVIWAYVWFFFFV	88
	orf129ng	ALRKVSLLYVTLFRGTPLFVQIVIWAYVWFFFFVILHTAFLGNAMROSRRVPDKGRWIAG	120

An ORF129ng nucleotide sequence <SEQ ID 841> was predicted to encode a protein having amino acid sequence <SEQ ID 842>:

5

1 MDRFRDIIYE YRWMFLYGAL TTLGLTVVAT AGGSVLGLLL ALARLIHLEK  
51 AGAPMRVLAW ALRKVSLLYV TLFRTPLFV QIVIWAYVWF PFFVILHTAF  
101 LGNAMRQSRP VPDKGRWIAG SLELNCQPRG RKTRGEFPPG ESNLGTEPRN  
151 PLMSGQRFP GCENWYPPQN FIKK\*

Further work revealed the following gonococcal sequence <SEQ ID 843>:

	1	ATGGATTTC	gtTTTGACAT	TATTTAcgaA	TACCGCTGGA	TGTTTTCTTTA
10	51	CGGCGCACTG	Acgaccttgg	ggctgacggt	cgtggcgacg	gCGGGCGGTT
	101	CGGtattggT	TCTGTTGTTG	GC GTTGGCGC	GCCTGATTCA	CTTGAAAACT
	151	CCCGGTGCGC	CGATGCGCGT	GCTGGCGTGG	CGGTGTGCGTA	AGGTTTCCGA
	201	GCTGTACGTT	ACCCTGTTCC	GGGGTACGCC	GCTGTTTG TG	CAGATTGTGA
15	251	TTTGGGCGTA	TGTGTGTTT	CCGTTTTTCG	TCCATCCTTT	AGACGGCATT
	301	TTGGTCAGCG	GCGAGGCGGC	AATCGCGCTG	CTCGCGGAT	ACGGGCCGCT
	351	GATTGCCGGT	TCTTTGGCAC	TGATCGCCAA	CTCGGGGGCG	TATATCTGTG
	401	AGATTTTCCG	CGCGGGCATC	CAGTCTATAG	ACAAAGGACA	GATGGAGGCG
20	451	GC GTGTCTT	TGGGACTGAC	CTATCCGCAG	CGGATGCGCT	ATGTTATTCT
	501	CCCGCAGGCA	TTGGCCGTA	TGCTGCCGCC	TTTGGCGAGC	GAGTTTCATCA
	551	CGCTCTTGAA	AGACAGCTCG	CTGCTGTCGG	TCATTGCTGT	GGCGGAGTTG
	601	GC GTATGTT	AGAATACGAT	TACGGGGCCG	TATTCGGTTT	ATGAAGAACC
	651	GCTTTACACC	GCCGCCCTGA	TTTATCTGTT	GATGACGATT	TTCTTAGGCT
	701	GGATATTCTT	CGGTTTGGAA	AAACGTTACA	ATCCGCAACA	CCGCTGA

This corresponds to the amino acid sequence <SEQ ID 844; ORF129ng-1>:

25

1	MDFRFDIIYE	YRWMFLYGAL	TTLGLTVVAT	AGGSVLGLLL	ALARLIHLEK
51	AGAPMRVLAW	ALRKVSLLYV	TLFRGTPLEFV	QIVIWAYVWF	FFFVHPSDGI
101	LVSGEAAIAL	RRGYPLIAG	SLALIANSQA	YICEIFRAGI	QSIDKGQMEA
151	ARSLGLTPQ	AMRYVILPQA	LRRMLPPLAS	EFITLLKDSS	LLSVIAVMAE
201	AYVONTITGR	YSVYEEPLYT	VALIYLLMTT	FLGWITFLRL	KRYNPQHR*

ORF129ng-1 and ORF129-1 show 99.2% identity in 248 aa overlap:

30	orf129-1.pep	MDFRFDIIIEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW 
	orf129ng-1	MDFRFDIIIEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW
35	orf129-1.pep	ALRKVSLLYVTLFRGTPLFVQIVIWAYVWFFFFVHPSDGILVSGEAAIALRRGYGPLIAG 
	orf129ng-1	ALRKVSLLYVTLFRGTPLFVQIVIWAYVWFFFFVHPSDGILVSGEAAIALRRGYGPLIAG
40	orf129-1.pep	SLALIANSYGAYICEIFRAGIQSIDKGQMEAAARSLGLTYPQAMRYVILPQALRRMLPPLAS 
	orf129ng-1	SLALIANSYGAYICEIFRAGIQSIDKGQMEAAACSLGLTYPQAMRYVILPQALRRMLPPLAS
45	orf129-1.pep	EFITLLKDSSLLSVIAVAELAYVQNTITGRYSVYEEPLYTVALIYLLMTTFLGWIFLRLE 
	orf129ng-1	EFITLLKDSSLLSVIAVAELAYVQNTITGRYSVYEEPLYTAAALIYLLMTTFLGWIFLRLE
	orf129-1.pep	KRYNPQHRX 
	orf129ng-1	KRYNPQHRX

In addition, ORF129ng-1 is homologous to an ABC transporter from *A.fulgidus*:

50 2650409(AE001090) glutamine ABC transporter, permease protein (glnP)  
[Archaeoglobus fulgidus]Length = 224  
Score = 132 bits (329), Expect = 2e-30  
Identities = 86/178 (48%), Positives = 103/178 (57%), Gaps = 18/178 (10%)

55 Query: 65 VSLLYVTLFRGTPLFVQIVIWAVVWFPPFVHPSDGLVSGEAAIALRRGYGPLIAGSLAL 124  
+S YV + RGTPL VQI+I +F P+ GI + E A G +AL  
Sbjct: 58 ISTATYVEVIRGTPLLVQILI-----VYFGLPAIGINLQPEA-----GIIAL 99



Query: 125 IANSGAYICEIFRAGIQSIDKGQMEAAACSLGLTYQAMRYVILPQALRRMLPPLASEFIT 184  
 SGAYI EI RAGI+SI QMEAA SLG+TY QAMRYVI PQA R +LP L +EFI  
 Sbjct: 100 SICSGAYIAEIVRAGIESIPIGQMEAAARSLGMTYLQAMRYVIFPQAFRNILPALGNEFIA 159

5 Query: 185 LLKDSSLLSVIAVAELAYVQNTITGRYSVYEEPLYTAALIYLLMTTFLGWIFLRLEKR 242  
 LLKDSSLLSVI++ EL V I P AL YL+MT L + +K+  
 Sbjct: 160 LLKDSSLLSVISIVELTRVGRQIVNTTFNAWTPFLGVALFYLMMTIPLSRLVAYSQKK 217

This analysis, including the identification of transmembrane domains in the two proteins, suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful  
 10 antigens for vaccines or diagnostics, or for raising antibodies.

### Example 100

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 845>:

1 ..CTGAAAGAAT GCCGTCTGAA AGACCCTGTT TTTATTCCAA ATATCGTTTA  
 51 TAAGAACATC GCCATTACTT TCCTGCTCTT GCACGCCGCC GCCGAACTTT  
 101 GGCTGCCCGC GCAAACCGCC GGTTTTACCG CGCTCGCCGT CGGCTTCATC  
 151 CTGCTCGCCA AGCTGCGTGA gCTTCACCAT CACGAACCT TACGTAAACA  
 201 cTACGTCCGC ACTTATTACy TGCTCCAAC TTTGCCGCC GCAGgcTAgt  
 251 TTGTGGACAG GCGCGCGwA ATTACAAAAC CTGCCCGCyT CCGCGCCCT  
 301 GCACCTGATT ACCCTCGGCG GCATGATGGG CGGCGTGATG ATGtGTGGc  
 351 TGACCGCCGG ACTGTGGCAC AGCGGCTTTA CCAAACCTCGA CTACCCCAA  
 401 CTCTGCCGCA TTGCCGTCCC CATCCTTTT GCGCGCGCG TCTCGCGCGC  
 451 TTTCTTGrTG AACGTGAACC CGrTATTTT CATTACCGTT CCTGCGATT  
 501 TGACCGCCGC CGTATTCGTA CTGTATCTT TCrCGTTTAT ACCGATATTT  
 551 CGGGCGAATG CGTTTACAGA CGATCCGGAr TA<sub>r</sub>

25 This corresponds to the amino acid sequence <SEQ ID 846; ORF130>:

1 ..LKECRLKDPV FIPNIVYKNI AITFLLHAA AELWLPAQTA GFTALAVGFI  
 51 LLAKLRELHH HELLRKHYVR TYYLLQLFAA AGSLWTGAAX LQNLPASAPL  
 101 HLITLGGMMG GVMVWLTA LWHSGFTKLD YPKLCRIAVP ILFAAAVSRA  
 151 FLXNVNPXFF ITVPAILTAA VFVLYLFXFI PIFRANAFTD DPE\*

30 Further work revealed the complete nucleotide sequence <SEQ ID 847>:

1 ATGCGGCCGT TTTTCGTCGG CGCGCGGGT CTTGCCATAC TCGGTGCGCT  
 51 GGTGTTTTTC ATCAACCCCG GTGCCATCGT CCTGCACCGC CAAATTTTCT  
 101 TGGAACTTAT GCTGCCGCG GCATACGCG GTTTTTGAC TCGCGCTTG  
 151 TTGGA CTGGA CGGGTTTTTC GGTAACTG AAACCTGTG CACTTTGAT  
 201 GCGGCGATTA TTGCTCGCG CATCCGCTAT ACTGCCCTTT TCGCGCAAA  
 251 CTGCTCGCT TTTCTGCGC GCCTATTGGC TGGTGTGCT GCTGTCTGC  
 301 GCGCGCTGA TTTGGCTAGA CCGAAACACC GACAACTCG CCCTGCTAAT  
 351 GTTACTTGCC GCGTTCACG TTTTTCAGAC GGCATATGCC GTCAGCGCG  
 401 ATTTGAACCT GTTGCGCGC CAAGTGCATC TAAATATGG GCGGTGATG  
 451 TTCGTATCCG TGCGCGTCAG TATTCTTTG GCGCGGAAG CCCTGAAAGA  
 501 ATGCCGTCTG AAAGACCCTG TTTTATTCC AAATATCGT TATAAAAACA  
 551 TCGCCATTAC TTTCTGCTC TTGCACGCG CCGCGAAT TTGGCTGCCC  
 601 GCGCAAACCG CCGGTTTAC CGCGCTCGC GTCGGCTTCA TCCTGCTCGC  
 651 CAAGCTGCGT GAGCTTCACC ATCAGCAACT CTTACGTAAC CACTACGTCC  
 701 GCACTTATTA CCTGCTCCAA CTCTTTGCC CCGCAGGCTA TTTGTGGACA  
 751 GCGCGGCGA AATTACAAA CCTGCCCGC TCCGCGCCC TGCACCTGAT  
 801 TACCTCGGC GGCATGATG GCGCGGTGAT GATGGTGTG CTGACCGCG  
 851 GACTGTGGCA CAGCGGCTT ACCAACTCG ACTACCCAA ACTCTGCCG  
 901 ATTGCCGTCC CCATCTTTT CCGCGCGCG GTCTCGCGC CTTTCTTGAT  
 951 AACGTGAAC CCGATATTT TCATTACCG TCCTGCGATT CTGACCGCG  
 1001 CCGTATTCG ACTGATCTT TCACGTTA TACCGATATT TCGGGCGAAT  
 1051 GCGTTTACAG ACGATCCGGA ATAA

This corresponds to the amino acid sequence <SEQ ID 848; ORF130-1>:

1 MRPFVGA AV LAILGALVFF INPGAIVLHR QIFLELMLPA AYGGFLTAAL  
 51 LDWTGFSGNL KPVATLMAAL LLAASAILPF SPQTASFFVA AYWLVLFFF

5

Homology with a predicted ORF from *N.meningitidis* (strain A)

10 *meningitidis:*

15

35

40

45

50

55

1 MRPFFVGA AV LAILGALVFF INPGAIVLHR QIFLELMLPA AYGGFLTAAL  
51 LDWTGFSGNL KPVATLMAAL LLAASAILPF SPOTASFFVA AYWLVLLLFC

5 101 ARLIWLD RNT DNFALLMLLA AFTVFQTAYA VSGDLNLLRA QVHLNMAAVM  
 151 FVSVRVSILL GAEALKECRL KDPVFIPNVV YKNIAITFLL LHAAELWLP  
 201 AQTAGFTSLA VGFI LLAKLR ELHHHELLRK HYVRTYYLLQ LFAAAGYLWT  
 251 GAAKLQNLPA SAPLHLITLG GMMGSVMVW LTAGLWHS GF TKLDYPKLCR  
 301 IAVPILFAAA VSRVLMNVN PIFFITVPAI LTAAVFVLYL LTFVPIFRAN  
 351 AFTDDPE\*

ORF130a and ORF130-1 show 98.3% identity in 357 aa overlap:

10 orf130a.pep MRPFVGA AVLAILGALVFFINPGAIVLHRQIFLEMLPAAYGGFLTAAALDWTGFSGNL  
 orf130-1 MRPFVGA AVLAILGALVFFINPGAIVLHRQIFLEMLPAAYGGFLTAAALDWTGFSGNL  
 15 orf130a.pep KPVATLMAALLLAASAILPFSPQTASFFVAAYWLVL LFCARLIWLD RNTDNFALLMLLA  
 orf130-1 KPVATLMAALLLAASAILPFSPQTASFFVAAYWLVL LFCARLIWLD RNTDNFALLMLLA  
 20 orf130a.pep AFTVFQTAYAVSGDLNLLRAQVHLNMAAVMFVSVRVSILLGAEALKECRLKDPVFIPNVV  
 orf130-1 AFTVFQTAYAVSGDLNLLRAQVHLNMAAVMFVSVRVSILLGAEALKECRLKDPVFIPNIV  
 25 orf130a.pep YKNIAITFLL LHAAELWLP AQTAGFTSLAVGFILLAKLRELHHHELLRKHYVRTYYLLQ  
 orf130-1 YKNIAITFLL LHAAELWLP AQTAGFTALAVGFILLAKLRELHHHELLRKHYVRTYYLLQ  
 30 orf130a.pep LFAAAGYLWTGA AKLQNLPA SAPLHLITLGMMGSVMVWLTAGLWHS GF TKLDYPKLCR  
 orf130-1 LFAAAGYLWTGA AKLQNLPA SAPLHLITLGMMGSVMVWLTAGLWHS GF TKLDYPKLCR  
 orf130a.pep IAVPILFAAAVSRVLMNVNPIFFITVPAILTA AVFVLYLLTFVPIFRANAFTDDPE  
 orf130-1 IAVPILFAAAVSRVLMNVNPIFFITVPAILTA AVFVLYLLTFVPIFRANAFTDDPE

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF130 shows 91.7% identity over a 193 aa overlap with a predicted ORF (ORF130ng) from *N.gonorrhoeae*:

35 orf130.pep LKECRLKDPVFIPNIVYKNIAITFLL LHAA 30  
 orf130ng LNLLRAQVHLNMAAVMFVSVRVS VLLGTETLKECRLKDPVFIPNVIYKNIAIT-LLHAA 201  
 40 orf130.pep AELWLPAQTAGFTALAVGFILLAKLRELHHHELLRKHYVRTYYLLQLFAAAGSLWTGAAX 90  
 orf130ng AELWLPAQTAGFTALAVGFILLAKLRELHHHELLRKHYVRTYYLLQLFAAAGYLWTGAAX 261  
 45 orf130.pep LQNLPA SAPLHLITLGMMGSVMVWLTAGLWHS GF TKLDYPKLCRIAVPILFAAAVSRA 150  
 orf130ng LQNLPA SAPLHLITLGGMTGGVMMVWLTAGLWHS GF TKLDYPKLCRIAVSILFASAVSRA 321  
 orf130.pep FLXNVNXPFFITVPAILTA AVFVLYLFXFIPIFRANAFTDDPE 193  
 orf130ng VLMNVNPIFFITVPEILTA AVFVLYLTFVPIFRANAFTDDPE 364

50 An ORF130ng nucleotide sequence <SEQ ID 851> was predicted to encode a protein having amino acid sequence <SEQ ID 852>:

55 1 MNKEFFTHPMR PFFVGA AVL I LGALVFFHQ PRRYHPAPPN FLGTYAAGCI  
 51 RRFEDYRFVG PDGFFRQPET CRYFDGGVVA CCGCFIAVFT ATCRIFRRRL  
 101 LAGVA AVLRL ADLARRQHRT LRSVDVTA AF TVFQTAYAVS GDLNLLRAQV  
 151 HLNMAAVMFV SVRVS VLLGT ETLKECRLKD PVFIPNVIYK NIAITLLLHA  
 201 AELWLPAQT AGFTALAVGF ILLAKLRELH HHELLRKHYV RTYYLLQLFA  
 251 AAGYLWTGAA KLQNLPA SAP LHLITLGGMT GGVMVWLTAGLWHS GF TKL  
 301 DYPKLCRIAV SILFASAVSR AVL MNVNPIF FITVPEILTA AVFVLYLLTF  
 351 VPIFRANAFT DDPE\*

Further work revealed the following gonococcal DNA sequence <SEQ ID 853>:

```

1  ATGCGCCCGT  TTTTCGTCGG  TCGGGCAGTA  CTTGCCATAC  TCGGTGCGTT
51  GGTGTTTTTT  ATCAACCCCG  GCGCTATCAT  CCTGCACCGC  CAAATTTTCT
101  TGGAACTTAT  GCTGCCGGCT  GCATACGGCG  GTTTTTTGAC  TACCGCTTTG
151  TTGGACCGGA  CGGGTTTTTC  AGGCAACCTG  AAACCTGCCG  CTACTTTGAT
201  GCGGGTGTGT  TTGCTTGTTG  CGGCTGTTTT  ATTGCCGTTT  TTACCGCAAC
251  TTGCGGCATT  TTTTCGTCGC  GCCTATTGGC  TGGTGTGTCT  GCTGTTCTGC
301  GCCTGGCTGA  TTTGGCTCGA  CCGCAACACC  GACAACTTCG  CTCTGTGTAT
351  GTTACTTGCC  GCATTACCG  TTTTTCAGAC  GGCCTATGCC  GTCAGCGGCG
401  ATTTGAACTT  ACTGCGCGCG  CAAGTGCATT  TGAATATGGC  GCGGTTCATG
451  TTCGTATCCG  TCCGCGTCAG  CGTCCTTTTG  GGCACGGAAA  CCCTGAAAGA
501  ATGCCGTCTG  AAAGACCCCG  TATTCATCCC  CAAGGTTATC  TATAAAAACA
551  TCGCCATCAC  CCTGCTGCTG  CACGCCGCGG  CCGAACTTTG  GCTGCCCGCG
601  CAAACCGCCG  GTTTTACTGC  GCTTGCCGTC  GGCTTCATCC  TGCTCGCCAA
651  CTGCGCGGAA  CTGCACATC  ACGAACTCTT  ACGCAAACAC  TACGTCCGCA
701  CTTATTACCT  GCTCCAGCTC  TTGCGCGCG  CAGGTTATCT  GTGGACAGGC
751  CCGCGGAAAC  TGCAAACCT  GCCCGCCTCC  GCGCCCCTGC  ACCTGATTAC
801  CCTCGCGGCG  ATGACGGGTG  GCGTGATGAT  GGTGTGGCTG  ACTGCCGGAC
851  TGTGGCACAG  CGGCTTTACC  AAACCTCGAC  ACCCGAAACT  CTGCCGCATC
901  GCGGTCTCCA  TCCTTTTCGC  CTCGCGCGGT  TCGCGCGCTG  TTTTAATGAA
951  CGTGAATCCG  ATATTCTTCA  TCACCGTTCC  CGAGATTCTG  ACCGCCCGCG
1001  TGTTCATGCT  TTACCTGCTG  ACGTTCGTAC  CGATTTTTCG  AGCGAACCGG
1051  TTTACAGACG  ATCCGGAATA  A

```

This corresponds to the amino acid sequence <SEQ ID 854; ORF130ng-1>:

```

1  MRPFFVGA AV LAILGALVFF INPGA IILHR QIFLELMLPA AYGGFLTTAL
51  LDRTGFSGNL KPAATLMAVL LLVAAVLLPF LPQLAAFFVA AYWLVLVLLFC
101  AWLIWLDNRNT DNFALLMLLA AFTVFQ TAYA VSGDLNLLRA QVHLNMAAVM
151  FVSVRVSVLL GTETLKECRL KDPVFIPNVI YKNIAITLLL HAAAEWLPA
201  QTAGFTALAV GFILLAKLRE LHHHELLRKH YVRTYLLQL FAAAGYLWTG
251  AAKLQNLPA S APLHLITLGG MTGGVMMVWL TAGLWHS GFT KLDYPKLCRI
301  AVSILFASAV SRAVLMNVNP IFFITVPEIL TAAVFMLYLL TFVPIFRANA
351  FTDDPE*

```

ORF130ng-1 and ORF130-1 show 92.4% identity in 357 aa overlap:

```

35  orf130-1.pep  MRPFFVGA AV LAILGALVFF INPGA IILHRQIFLELMLPAAYGGFLTTALLDWTGFSGNL
    orf130ng-1  MRPFFVGA AV LAILGALVFF INPGA IILHRQIFLELMLPAAYGGFLTTALLDRTGFSGNL

40  orf130-1.pep  KPVATLMAALLAASAILPFSPQTASFFVAAYWLVLVLLFCARLIWLDNRNTDNFALLMLLA
    orf130ng-1  KPAATLMAVLLLVAAVLLPFLPQLAAFFVAAYWLVLVLLFCAWLIWLDNRNTDNFALLMLLA

45  orf130-1.pep  AFTVFQ TAYAVSGDLNLLRAQVHLNMAAVMFVSVRVSI LLGAEALKECRLKDPVFIPNIV
    orf130ng-1  AFTVFQ TAYAVSGDLNLLRAQVHLNMAAVMFVSVRVSVLLGTETLKECRLKDPVFIPNVI

50  orf130-1.pep  YKNIAITFLLLHAAAEWLPAQTAGFTALAVGFILLAKLRELHHHELLRKHYVRTYLLQ
    orf130ng-1  YKNIAIT-LLLHAAAEWLPAQTAGFTALAVGFILLAKLRELHHHELLRKHYVRTYLLQ

55  orf130-1.pep  LFAAGYLWTGA AKLQNLPA S APLHLITLGGMMGGVMMVWL TAGLWHS GFT KLDYPKLCR
    orf130ng-1  LFAAGYLWTGA AKLQNLPA S APLHLITLGGMTGGVMMVWL TAGLWHS GFT KLDYPKLCR

    orf130-1.pep  IAVPILFAAAVSRAFLMNVNPIFFITVPAILTAAVFVLYLFTFIPIFRANAFTDDPEX
    orf130ng-1  IAVSILFASAVSRAVLMNVNPIFFITVPEILTAAVFMLYLLTFVPIFRANAFTDDPEX

```

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

**Example 101**

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 855>:

```

1  ATGGAATTC GGGCAATAAA ATATACGGCA ATGGCTGCGT TGCTTGCAAT
5  51  TACGGTTGCA GGCTGCCGGC TGGCGGGGTG GTATGAGTGT TCGTCCCTCA
    101 CCGGCTGGTG TAAGCCGAGA AAACCGGCTG CCATCGATTT TTGGGATATT
    151 GCGGCGGAGA GTCCGCCGTC TTTAGGGGAC TACGAGATAC CGCTTTCAGA
    201 CGGCAATAGT TCCGTCAGGG CAAACGAATA TGAATCCGCA CAACAATCTT
    251 ACTTTTACAG GAAAATAGGG AAGTTTGAAG C.TGCGGGCT GGATTGGCGT
    301 ACGCGTGACG GCAAACCTTT GATTGAGACG TTCAAACAGG GAGGATTTGA
10  351 CTGCTTGGA AAG..

```

This corresponds to the amino acid sequence <SEQ ID 856; ORF131>:

```

1  MEIRAIKYTA MAALLAFTVA GCRLAGWYEC SSLTGWCKPR KPAAIDFWDI
51  GGESPPSLGD YEIPLSDGNS SVRANEYESA QSYFYRKIG KFEXCGLDWR
101 TRDGKPLIET FKQGGFDCLE K..

```

Further work revealed the complete nucleotide sequence <SEQ ID 857>:

```

1  ATGGAATTC GGGCAATAAA ATATACGGCA ATGGCTGCGT TGCTTGCAAT
20  51  TACGGTTGCA GGCTGCCGGC TGGCGGGGTG GTATGAGTGT TCGTCCCTCA
    101 CCGGCTGGTG TAAGCCGAGA AAACCGGCTG CCATCGATTT TTGGGATATT
    151 GCGGCGGAGA GTCCGCCGTC TTTAGGGGAC TACGAGATAC CGCTTTCAGA
    201 CGGCAATCGT TCCGTCAGGG CAAACGAATA TGAATCCGCA CAACAATCTT
    251 ACTTTTACAG GAAAATAGGG AAGTTTGAAG CCTGCGGGCT GGATTGGCGT
    301 ACGCGTGACG GCAAACCTTT GATTGAGACG TTCAAACAGG GAGGATTTGA
    351 CTGCTTGGA AAGCAGGGGT TGGCGCGCAA CGGTCTGTCC GAGCGCGTCC
    401 GATGGTAA

```

This corresponds to the amino acid sequence <SEQ ID 858; ORF131-1>:

```

1  MEIRAIKYTA MAALLAFTVA GCRLAGWYEC SSLTGWCKPR KPAAIDFWDI
51  GGESPPSLGD YEIPLSDGNS SVRANEYESA QSYFYRKIG KFEACGLDWR
101 TRDGKPLIET FKQGGFDCLE KQGLRRNGLS ERVRW*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF131 shows 95.0% identity over a 121aa overlap with an ORF (ORF131a) from strain A of *N.meningitidis*:

```

35  orf131.pep  10      20      30      40      50      60
    MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPPSLGD
    orf131a    10      20      30      40      50      60
    MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLSGWCKPRKPAAIDFWDIGGESPPSLED

40  orf131.pep  70      80      90      100     110     120
    YEIPLSDGNSSVRANEYESAQSYFYRKIGKFEXCGLDWRTRDGKPLIETFKQGGFDCLE
    orf131a    70      80      90      100     110     120
    YEIPLSDGNRSVRANEYESAQSYFYRKIGKFEXCGLDWRTRDGKPLIETFKQEGFDCLK

45  orf131.pep  K
    |
    orf131a    KQGLRRNGLSERVRWX
                130

```

The complete length ORF131a nucleotide sequence <SEQ ID 859> is:

```

1  ATGGAATTC GGGCAATAAA ATATACGGCA ATGGCTGCGT TGCTTGCAAT
51  TACGGTTGCA GGCTGCCGGT TGGCAGGTTG GTATGAGTGT TCGTCCCTGT

```

5  
101 CCGGCTGGTG TAAGCCGAGA AAACCTGCCG CCATCGATTT TTGGGATATT  
151 GCGGCGGAGA GTCCTCCGTC TTTAGAGGAC TACGAGATAC CGCTTTCAGA  
201 CGGCAATCGT TCCGTCAGGG CAAACGAATA TGAATCCGCA CAACAATCTT  
251 ACTTTTACAG GAAAATAGGG AAGTTTGAAG CCTGCGGGTT GGATTGGCGT  
301 ACGCGTGACG GCAAACCTTT GATTGAGACG TTCAAACAGG AAGGTTTGA  
351 TTGTTTGAAA AAGCAGGGGT TGCGGCGCAA CGGTCTGTCC GAGCGCGTCC  
401 GATGGTAA

This encodes a protein having amino acid sequence <SEQ ID 860>:

10  
1 MEIRAIKYTA MAALLAFTVA GCRLAGWYEC SSLSGWCKPR KPAAIDFWDI  
51 GGESPPSLED YEIPLSDGNR SVRANEYESA QSYFYRKIG KFEACGLDWR  
101 TRDGKPLIET FKQEGFDCLK KQGLRRNGLS ERVRW\*

ORF131a and ORF131-1 show 97.0% identity in 135 aa overlap:

15  
orf131a.pep MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLSGWCKPRKPAAIDFWDIGGESPPSLED  
orf131-1 MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPPSLGD  
orf131a.pep YEIPLSDGNRSVRANEYESAQSYFYRKIGKFEACGLDWRTRDGKPLIETFKQEGFDCLK  
20 orf131-1 YEIPLSDGNRSVRANEYESAQSYFYRKIGKFEACGLDWRTRDGKPLIETFKQGGFDCLE  
orf131a.pep KQGLRRNGLSERVRWX  
orf131-1 KQGLRRNGLSERVRWX

## 25 Homology with a predicted ORF from *N.gonorrhoeae*

ORF131 shows 89.3% identity over 121 aa overlap with a predicted ORF (ORF131ng) from *N.gonorrhoeae*:

30  
orf131.pep MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPPSLGD 60  
orf131ng MEIRVIKYTATAALFAFTVAGCRLAGWYECSSLSGWCKPRKPAAIDFWDIGGESPLSLED 60  
orf131.pep YEIPLSDGNSSVRANEYESAQSYFYRKIGKFEXCGLDWRTRDGKPLIETFKQGGFDCLE 120  
orf131ng YEIPLSDGNRSVRANEYESAQSYFYRKIGKFEACGLDWRTRDGKPLVERFKQEGFDCLE 120  
35  
orf131.pep K 121  
orf131ng KQGLRRNGLSERVRW 134

40  
A complete length ORF131ng nucleotide sequence <SEQ ID 861> was predicted to encode a protein having amino acid sequence <SEQ ID 862>:

1 MEIRVIKYTA TAALFAFTVA GCRLAGWYEC LSLSGWCKPR KPAAIDFWDI  
51 GGESPLSLED YEIPLSDGNR SVRANEYESA QKSYFYRKIG KFEACGLDWR  
101 TRDGKPLVER FKQEGFDCLE KQGLRRNGLS ERVRW\*

Further work revealed the following gonococcal DNA sequence <SEQ ID 863>:

45  
1 ATGGAAATTC GGGTAATAAA ATATACGGCA ACGGCTGCGT TGTTTGCAAT  
51 TACGGTTGCA GGCTGCCGGC TGGCGGGGTG GTATGAGTGT TCGTCCCTGT  
101 CCGGCTGGTG TAAGCCGAGA AAACCTGCCG CCATCGATTT TTGGGATATT  
151 GCGGCGGAGA GtcgcgtGTC TTTAGAGGAC TACGAGATAC CGCTTTCAGA  
50 201 CGGCAATCGT TCCGTCAGGG CAAACGAATA TGAATCCGCG CAAAAATCTT  
251 ACTTTTATAG GAAAATAGGG AAGTTTGAAG CCTGCGGGTT GGATTGGCGT  
301 ACGCGTGACG GCAAACCTTT GGTGAGAGG TTCAAACAGG AAGGTTTCGA  
351 CTGTTTGAA AAGCAGGGGT TGCGGCGCAA CGGCCTGTCC GAGCGCGTCC  
401 GATGGTAA

This corresponds to the amino acid sequence <SEQ ID 864; ORF131ng-1>:

```

1 MEIRVIKYTA TAALFAFTVA GCRLAGWYEC SSLSGWCKPR KPAAIDFWDI
51 GGESPLSLED YEIPLSDGNR SVRANEYESA QKSYFYRKIG KFEACGLDWR
101 TRDGKPLVER FKQEGFDCLE KQGLRRNGLS ERVRW*

```

ORF131ng-1 and ORF131-1 show 92.6% identity in 135 aa overlap:

```

5 orf131ng-1.pep MEIRVIKYTATAALFAFTVAGCRLAGWYECSSLSGWCKPRKPAAIDFWDIGGESPLSLED
   orf131-1      MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPPSLGD

   orf131ng-1.pep YEIPLSDGNRSVRANEYESAQKSYFYRKIGKFEACGLDWRTRDGKPLVERFKQEGFDCLE
10 orf131-1      YEIPLSDGNRSVRANEYESAQKSYFYRKIGKFEACGLDWRTRDGKPLIETFKQGGFDCLE

   orf131ng-1.pep KQGLRRNGLSERVRWX
   orf131-1      KQGLRRNGLSERVRWX
15

```

Based on the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 102

20 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 865>

```

1 ATGAAACACA TCCATATTAT CGGTATCGGC GGCACGTTTA TGGGCGGGCT
51 TGCCGCCATT GCCAAAGAAG CGGGGTTTGA AGTCAGCGGT TGCGACGCGA
101 AGATGTATCC GCCGATGAGC ACCCAGCTCG AAGCCTTGGG TATAGACGTG
25 151 TATGAAGGCT TCGATGCCGC TCAGTTGGAC GAATTTAAAG CCGACGTTTA
   201 CGTTATCGGC AATGTCGCCA AGCGCGGGAT GGATGTGGTT GAAGCGATTT
   251 TGAACCTCGG CCTGCCTTAT ATtTcCGGCC CGCAATGGCT GTCGGAAAAC
   301 GTGCTGCACC ATCATTGGGT ACTCGGTGTG GCGGGGACgC ACGGCAAAAC
   351 GACCACCGCC TCCATGCTCG CATGGGTCTT GGAATATgCC GGCCTCGCGC
   401 CGGGCTTCCT TATtGGCGGC GTACC.GGAA AATtCGGCC TTTCCGCCCG
30 451 CCTGCCGCAA ACGCCGCGCC AAGACCCGAA CAGCCAATCG CCGTTTTTcG
   501 TCATCGAAGC CGACGAATAC GACACCGCCT TtTCGACAA ACgTTCTAAA
   551 TtCGTGcATT ACCGTCCGCG TACCGCCGTG TTGAACAATC TGGAATTcGA
   601 CCACGCCGAC ATCTTTGCCG ACTTGGGCGC GATACAGACC CAGTTCCACT
   651 ACCTCGTGCG TACCGTGCCG TCTGAAGGCT TAATCGTCTG CAACGGACGG
35 701 CAGCAAAGCC TGCAAGATAC TTTGGACAAA GGCTGCTGGA CGCCGGTGGa
   751 AAAATTCCGC ACGGAACACG GCTGGCA..

```

This corresponds to the amino acid sequence <SEQ ID 866; ORF132>:

```

1 MKHIHIIGIG GTFMGGLAAl AKEAGFEVSG CDAKMYPPMS TQLEALGIDV
51 YEGFDAAQLD EFKADVYVIG NVAKRGMdVv EAILNLGLPY ISGPQWLSEN
40 101 VLHHHWVLGV AGTHGKTTTA SMLAWVLEYA GLAPGFLIGG VXGKFRFRFP
   151 PAANAAPRPE QPIAVERHRS RRIRHRLFRQ TFXIRALPSA YRRVEQSGIR
   201 PRRHLCLRGR DTDVPLPRa YRAVXLNRL QRTAAKPARY FGQRLLDAGG
   251 KIRHGTRLA..

```

Further work revealed the complete nucleotide sequence <SEQ ID 867>:

```

45 1 ATGAAACACA TCCATATTAT CGGTATCGGC GGCACGTTTA TGGGCGGGCT
   51 TGCCGCCATT GCCAAAGAAG CGGGGTTTGA AGTCAGCGGT TGCGACGCGA
   101 AGATGTATCC GCCGATGAGC ACCCAGCTCG AAGCCTTGGG TATAGACGTG
   151 TATGAAGGCT TCGATGCCGC TCAGTTGGAC GAATTTAAAG CCGACGTTTA
50 201 CGTTATCGGC AATGTCGCCA AGCGCGGGAT GGATGTGGTT GAAGCGATTT
   251 TGAACCTCGG CCTGCCTTAT ATTTCCGCCC CGCAATGGCT GTCGGAAAAC
   301 GTGCTGCACC ATCATTGGGT ACTCGGTGTG GCGGGGACGC ACGGCAAAAC
   351 GACCACCGCC TCCATGCTCG CATGGGTCTT GGAATATGCC GGCCTCGCGC
   401 CGGGCTTCCT TATtGGCGGC GTACCGGAAA ATTTCGGCGT TTCCGCCCGC
   451 CTGCCGCAA CGCCGCGCCA AGACCCGAAC AGCCAATCGC CGTTTTTCGT
55 501 CATCGAAGCC GACGAATACG ACACCGCCTT TTTCGACAAA CGTTCTAAAT

```

551 TCGTGCATTA CCGTCCGCGT ACCGCCGTGT TGAACAATCT GGAATTCGAC  
 601 CACGCCGACA TCTTTGCCGA CTTGGGCGCG ATACAGACCC AGTTCACCTA  
 651 CCTCGTGCGT ACCGTGCCGT CTGAAGGCTT AATCGTCTGC AACGGACGGC  
 701 AGCAAAGCCT GCAAGATACT TTGGACAAAG GCTGCTGGAC GCCGTGGAA  
 751 AAATTCGGCA CGGAACACGG CTGGCAGGCC GCGGAAGCCA ATGCCGACGG  
 801 CTCGTTCGAC GTGTTGCTCG ACGGCAAAAC CGCCGGACGC GTCAAATGGG  
 851 ATTTGATGGG CAGGCACAAC CGCATGAACG CGCTCGCCGT CATTGCCGCC  
 901 GCGCGTCATG TCGGTGTCGA TATTCAGACC GCCTGCGAAG CCTTGGGCGC  
 951 GTTTAAAAAC GTCAAACGCC GGATGGAAAT CAAAGGCACG GCAAACGGCA  
 1001 TCACCGTTTA CGACGACTTC GCCCACCACC CGACCGCCAT CGAAACACG  
 1051 ATTCAAGGTT TGCGCCAACG CGTCGGCGCG GCGCGCATCC TCGCCGTCTT  
 1101 CGAACC GCGT TCCAACACGA TGAAGCTGGG CACGATGAAG TCCGCCCTGC  
 1151 CTGTAAGCCT CAAAGAGGCC GACCAAGTGT TCTGCTACGC CGGCGGCGTG  
 1201 GACTGGGACG TCGCCGAAGC CCTCGCGCCT TTGGGCGGCA GGCTGAACGT  
 1251 CGGCAAAGAC TTCGATGCCT TCGTTGCCGA AATCGTGAAG AACGCCGAAG  
 1301 TAGGCGACCA TATTTTGGTG ATGAGCAACG GCGGTTTCGG CGGAATACAC  
 1351 GGAAAGCTGC TGAAGCTTT GAGATAG

This corresponds to the amino acid sequence <SEQ ID 868; ORF132-1>:

1 MKHIHIIGIG GTFMGGGLAAI AKEAGFEVSG CDAKMYPPMS TQLEALGIDV  
 51 YEGFDAAQLD EFKADVIVIG NVAKRGMV V EAILNLGLPY ISGPQWLSN  
 101 VLHHHWVLGV AGTHGKTTTA SMLAWVLEYA GLAPGFLIGG VPENFGVSAR  
 151 LPQTPRQDPN SQSPFFVIEA DEYDTAFFDK RSKFVHYRPR TAVLNNLEFD  
 201 HADIFADLGA IQTQFHYLVR TVPSEGLIVC NGRQSLQDT LDKGCWTPVE  
 251 KFGTEHGWQA GEANADGSFD VLLDGKTAGR VKWDLMGRHN RMNALAVIAA  
 301 ARHVGVDIQT ACEALGAFKN VKRRMEIKGT ANGITYYDDF AHHPTAIETT  
 351 IQGLRQRVGG ARILAVLEPR SNTMKLGTMK SALPVSLKEA DQVFCYAGGV  
 401 DWDAEALAP LGRLNVGKD FDAFVAEIVK NAEVGDHILV MSNGGFGGIH  
 451 GKLEALR\*

Computer analysis of this amino acid sequence gave the following results:

30 Homology with the hypothetical o457 protein of *E.coli* (accession number U14003)

ORF132 and o457 show 58% aa identity in 140 aa overlap:

Orf132: 4 IHIIGIGGTFMGGGLAAIAKEAGFEVSGCDAKMYPPMSTQLEALGIDVYEGFDAAQLDEFK 63  
 IHI+GI GTFMGGGLA +A++ G EV+G DA +YPPMST LE GI++ +G+DA+QL+ +  
 o457: 3 IHILGICGTFMGGGLAMLARQLGHEVTGSDANVYPPMSTLLEKQGIELIQYDASQLEP-Q 61  
 Orf132: 64 ADVYVIGNVAKRGMVDVVEAILNLGLPYISGPQWLSENVLHHHWVLGVAGTHGKTTTASML 123  
 D+ +IGN RG VEA+L +PY+SGPQWL + VL WV L VAGTHGKTTTA M  
 o457: 62 PDLVIIGNAMTRGNPCVEAVLEKNIPYMSGPQWLHDFVLRDRVWLAVAGTHGKTTTAGMA 121  
 Orf132: 124 AWWLEYAGLAPGFLIGGVXG 143  
 W+LE G PGF+IGGV G  
 o457: 122 TWILEQCGYKPGFVIGGVPG 141

Homology with a predicted ORF from *N.meningitidis* (strain A)

45 ORF132 shows 74.6% identity over a 189aa overlap with an ORF (ORF132a) from strain A of *N. meningitidis*:

or132.pep MKHIHIIGIGGTFMGGGLAAIAKEAGFEVSGCDAKMYPPMSTQLEALGIDVYEGFDAAQLD  
 or132a MKHIHIIGIGGTFMGGGLAAIAKEAGFEVSGCDAKMYPPMSTQLEALGIDVYEGFDAAQLD  
 or132.pep EFKADVIVIGNVAKRGMVDVVEAILNLGLPYISGPQWLSENVLHHHWVLGVAGTHGKTTTA  
 or132a EFKADVIVIGNVAKRGMVDVVEAILNLGLPYISGPQWLAENXLLHHHWXLVAGTHGKTTTA



```

    orf132.pep      SMLAWVLEYAGLAPGFLIGGVXGKFR---RFRPPAANAAPRPEQPI-----AVFR
                    ||| ||||| ||||| |||| | : | : | : ||| | ||
    orf132a         SMLAWVLEYAGLAPGFIXIGGVPENFSVSARL-PQTPRQDPNSQSPFFVIEADEYDTAFFD
                    130       140       150       160       170

5
    orf132.pep      170       180       190       200       210       220
                    HRSRRIRHRLRFQTFXIRALPSAYRRVEQSGIRPRRHLCRLGRDTPVPLPRAYRAVXRL
                    :||: ::|
    orf132a         KRSKFVHYRPRTAVLNLEFDHADIFADLGAIQTQFHHLVRTVPSEGLIVCNGRQQSLQD
                    180       190       200       210       220       230

```

The complete length ORF132a nucleotide sequence <SEQ ID 869> is:

	1	ATGAAACACA	TCCACATTAT	CGGTATCGGC	GGCACGTTTA	TGGGTGGGAT
	51	TGCCGCCATT	GCCAAAGAAG	CAGGGTTTGA	ANTCAGCGGT	TGCGATGCGA
15	101	AGATGTATCC	GCCGATGAGC	ACCCAGCTCG	AAGCCTTGGG	CATAGGCGGT
	151	TATGAAGCTT	TCGACACCGC	GCAGTTGGAC	GAATTTAAAG	CCGACGTTTA
	201	CGTTATCGCG	AATGTCGCCA	AGCGCGGGAT	GGATGTGGTT	GAAGCGATTT
	251	TGAACCGTGG	GCTGCCTTAT	ATTTCGCGCC	CGCAATGGCT	GGCTGAAAAC
	301	NTGCTGCACC	ATCATTTGNN	ACTCGGCGTG	CGCGNGACGC	ACGGCAAAAC
20	351	GACCAACGCG	TCTATGTCG	CGTGGGTTTT	GGAAATATGG	GGACTCGCAC
	401	CGGGCTTCNT	TATCGGCGGC	GTACCGGAAA	ACTTCAGCGT	TTCCGCCCCG
	451	CTGCCGCAAA	CGCCGCGCCA	AGACCCGAAC	AGCCATATCG	CGTTTTTTCGT
	501	CATTGAAGCC	GACGAATACG	ACACCCGCTT	TTTCGACAAA	CGCTCCAAAT
	551	TCGTGCATTA	CCGTCCGCGT	ACCGCCGTGT	TGAAACAATT	GCAATTCGCAC
25	601	CACGCCGACA	TCTTCGCCGA	TTTGGGCGCG	ATACAGACCC	AGTTCACCA
	651	CCTCGTGCCT	ACCGTGCCGT	CTGAAGGCCT	CATCGTCTGC	AACGGACGGC
	701	AGCAAAGCCT	GCAAGACACT	TTGGACAAAG	GCTGCTGGAC	GCCGCTGGAA
	751	AAATTCCGCA	CGGAACACGG	CTGGCAGGCC	GCGCGAAGCC	ATGCCGATGG
	801	CTCGTTTCGAC	GTGTTGCTTG	ACGGCAAAAA	AGCCGGACAC	GTGCTTTGGA
30	851	GTTTTGATGG	CGGACACAAC	CGCATGAACG	CGCTCGCNGT	CATCGCCGCC
	901	GCGGTCATG	CCGGAGTNGA	CATTTCAGAC	GCCTGCGAAG	CCTTGAGCAC
	951	GTTTAAAAAC	GTCAAACGCC	GCATGGAAAT	CAAAAGCGAC	GCAAAACGGTA
	1001	TCACCGTTTA	CGACGACTTC	GCCCACCATC	CGACCGCTAT	CGAAACCACG
	1051	ATTCAAGGTT	TGCGCCAGCG	CGTCGCGCGC	GCGCGCATAT	TCGCCGTCCT
35	1101	CGAACC CGCT	TCCAATACGA	TGAAGCTGGG	TACGATGAAA	GCCGCCCTGC
	1151	CCGCAAGCCT	CAAGAAGACC	GACCAAGTGT	TCTGNTACG	GCGCGCGCGC
	1201	GACTGGGACG	TTGCCGAAGC	CCTCGCGCCT	TTGGGCGGCA	GGCTGCACGT
	1251	CGGCAAAAGC	TTTCGATGCC	TCGTTGCCGA	AATCGTGAAA	AACGCCGAAG
	1301	CAGGCGACCA	TATTTTGGTG	ATGAGCAACG	GCGGTTTCGG	CGGAATACAC
	1351	ACCAAACCTG	TGGACGCTTT	GAGATAG		

40 This encodes a protein having amino acid sequence <SEQ ID 870>:

	1	MKHIHIIGIG	GTFMGGIAAI	AKEAGFEXSG	CDAKMYPPMS	TQLEALGIGV
	51	YEGFDTAQLD	EFKADVYVIG	NVAKRGMDVV	EAILNRGLPY	ISGPQWLAE
	101	XLHHHWXLGV	AXTHGKTSTA	SLAWVLEYA	GLAPGFXIGG	VPENFVSAR
45	151	LPQTPRQDPN	QSPPFFVIEA	DEYDTAFFDK	RSKFVHYRPR	TAVLNNLEFD
	201	HADIFADLGA	IQTQFHHLVR	TVPSEGLIVC	NGRQSLQDTH	LDKGCTWPVE
	251	KFGTEHGWQA	GEANADGSFD	VLLDGKKAGH	VAWSLMGGHN	RMNALAVIAA
	301	ARHAGVDIQT	ACEALSTFKN	VKRMEIKGT	ANGITVYDDF	AHHPTAIETT
	351	IQGLRQRVGG	ARILAVLEPR	SNTMKLGTMK	AALPASLKEA	DQVFXYAGGA
50	401	DWDVAEALAP	LGGRHLVHGK	FDAFVAEIVK	NAEAGDHILV	MSNGGFGGIH
	451	TKLLDALR*				

ORF132a and ORF132-1 show 93.9% identity in 458 aa overlap:

	orf132a.pep	MKHIHIIGIGGTFMGGIAAIAKEAGFEXSGCDAKMYPPMSTQLEALGIGVYEGFDTAQLD
		:
55	orf132-1	MKHIHIIGIGGTFMGGIAAIAKEAGFEVSGCDAKMYPPMSTQLEALGIDVYEGFDAAQLD
	orf132a.pep	EFKADVYVIGNVAKRGMDVVEAILNRGLPYISGPQWLAE NXLHHHWXLGVAXTHGKTTTA
		:
	orf132-1	EFKADVYVIGNVAKRGMDVVEAILNLGLPYISGPQWLSENVLHHHWLVGVAGTHGKTTTA
60	orf132a.pep	SMLAWVLEYAGLAPGFXIGGVENFVS SARLPQTPRODPNSQSPFFVIEADEYDTAFFDK
		:
	orf132-1	SMLAWVLEYAGLAPGFLIGGVENFGVSARLPQTPRODPNSQSPFFVIEADEYDTAFFDK

	orf132a.pep	RSKFVHYRPTAVLNNLEFDHADIFADLGAIQTQFHHLVVRTVPSEGLIVCNGRQQSLQDT
	orf132-1	RSKFVHYRPTAVLNNLEFDHADIFADLGAIQTQFHHLVVRTVPSEGLIVCNGRQQSLQDT
5	orf132a.pep	LDKGCWTPVEKFGTEHGWQAGEANADGSFDVLLDGKKAGHVAWSLMGGHNRMNALAVIAA
	orf132-1	LDKGCWTPVEKFGTEHGWQAGEANADGSFDVLLDGKTAGRVKWDLMGRHNRMNALAVIAA
10	orf132a.pep	ARHAGVDIQTACEALSTFKNVKRRMEIKGTANGITVYDDFAHHPTAIETTIQGLRQRVGG
	orf132-1	ARHVGVDIQTACEALGAFKNVKKRMEIKGTANGITVYDDFAHHPTAIETTIQGLRQRVGG
	orf132a.pep	ARILAVLEPRSNMTMKLGTMKAAALPASLKEADQVFXYAGGADWDVAEALAPLGGRLHVGKD
15	orf132-1	ARILAVLEPRSNMTMKLGTMKALPVSLSLKEADQVFCYAGGVWDVAEALAPLGGRLNVGKD
	orf132a.pep	FDAFVAEIVKNAEAGDHILVMSNGGFGGIHTKLLDALRX
20	orf132-1	FDAFVAEIVKNAEAGDHILVMSNGGFGGIHGLLEALRX

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF132 shows 89.6% identity over 259 aa overlap with a predicted ORF (ORF132ng) from *N.gonorrhoeae*:

25	orf132.pep	MKHIHIIGIGGTFMGGIAAIAKEAGFEVSGCDKMYPPMSTQLEALGIDVYEGFDAAQLD	60
	orf132ng	MKHIHIIGIGGTFMGGIAAIAKEAGFKVSGCDKMYPPMSTQLEALGIGVHEGFDAQAQLE	60
30	orf132.pep	EFKADVYVIGNVAKRGMDVVEAILNLGLPYISGPQWLSENVLHHHWLVGVAGTHGKTTTA	120
	orf132ng	EFQADIYVIGNVARRGMDVVEAILNRGLPYISGPQWLAENVLHHHWLVGVAGTHGKTTTA	120
35	orf132.pep	SMLAWVLEYAGLAPGFLIGGVXGKFRFRPPAANAAPRPEQPIAVFRHRSRRIRHRLFRQ	180
	orf132ng	SMLAWVLEYAGLAPGFLIGGVPGKFRFRPPTANAASRPEQQIAVFRHRSRRIRHRLFRQ	180
40	orf132.pep	TFXIRALPSAYRRVEQSGIRPRRHLRLGRDTPVPLPRAYRAVXRLNRLQRTAAKPARY	240
	orf132ng	TLQIRALSPAYRRVEQSGIRPRRHLRLGRDTPVPPRAHRTIRRPRLQRTAAKPARY	240
	orf132.pep	FGQRLLDAGGKIRHGTRLA	259
	orf132ng	FGQRLLDAGGKIRHRLADW	261

An ORF132ng nucleotide sequence <SEQ ID 871> was predicted to encode a protein having amino acid sequence <SEQ ID 872>:

45	1	MKHIHIIGIG	GTFMGGIAAI	AKEAGFKVSG	CDKMYPPMS	TQLEALGIGV
	51	HEGFDAQAQLE	EFQADIYVIG	NVARRGMDVV	EAILNRGLPY	ISGPQWLAEN
	101	VLHHHWLVGV	AGTHGKTTTA	SMLAWVLEYA	GLAPGFLIGG	VPGKFRFRFP
	151	PTANAASRPE	QQIAVFRHRS	RRIRHRLFRQ	TLQIRALSPA	YRRVEQSGIR
50	201	PRRHLRLGR	DTDPVPPRA	HRTIRRPRL	QRTAAKPARY	FGQRLLDAGG
	251	KIRHRLAD	W*			

Further work revealed the following gonococcal DNA sequence <SEQ ID 873>:

	1	ATGAAACACA	TCCACATTAT	CGGTATCGGC	GGCACGTTTA	TGGGCGGGAT
	51	TGCCGCCATT	GCCAAAGAAG	CCGGGTTCAA	AGTCAGCGGT	TGCGACGCGA
55	101	AGATGTATCC	GCCGATGAGC	ACCCAGCTCG	AAGCCTTGGG	CATAGGCGTA
	151	CACGAAGGCT	TCCGATGCCG	GCAGTTGGAA	GAATTTCAAG	CCGATATTTA
	201	CGTCATCGGC	AATGTCGCCA	GGCGCGGGAT	GGATGTGGTC	GAGGCGATTT
	251	TGAACCGTGG	GCTGCCTTAT	ATTTCCGCC	CGCAATGGCT	GGCTGAAAc
	301	GTGctgcacc	atcaTTGGgt	ACTCGGCGTG	GcagggaCGC	ACGGcaaAac
60	351	gaccaCcGcg	tCCATGCTCG	CCTGGGTCTT	GGAATATGCC	GGACTCGCGC
	401	CGGGCTTCCT	CATCGGCGGT	gtaccggaAA	ATTTCCGGCT	TTCCGCCCGC

5	451	CTACCGCAAA	CGCCGCGTCA	AGACCCGAAC	AGCAAATCGC	CGTTTTTCGT
	501	CATCGAAGCC	GACGAATACG	ACACCGCCTT	TTTCGACAAA	CGCTCCAAAT
	551	TCGTGCATTA	TGCCCCGCGT	ACCCGCGTGT	TGAACAATCT	GGAATTTCCGAC
	601	CACGCCGACA	TCTTCGCGGA	CTTGGGCGCG	ATACAGACCC	AGTTTCCACCA
	651	CCTCGTGCGC	ACCGTACCAT	CCGAAGGCCT	CATCGTCTGC	AACGGACAGC
10	701	AGCAAAGCCT	GCAAGATACT	TTGGACAAAG	GCTGCTGGAC	GCCGCTGGAA
	751	AAATTTCGGCA	CCGGACACGG	CTGGCAGATT	GGTGAAGTCA	ATGCCGACGG
	801	CTCGTTTCGAC	GTATTCCTTG	ACGGCAAAAA	AGCCGACAC	GTCCGATGGG
	851	ATTTGATGGG	CGGACACAAC	CGCATGAACG	CGCTCGCCGT	CATCGTGGC
	901	GCACGCCATG	CCGGAGTCGA	TGTTTCAGACG	GCCCTGCGAAG	CCTTGGGTGC
15	951	GTTTAAAAAC	GTCAAACGCC	GCATGGAAT	CAAAAGCAGC	CGAAACGGCA
	1001	TCACCGTTTA	CGACGATTTC	GCCCAACACC	CGACCGCCAT	CGAAACCACG
	1051	ATTCAAGGTT	TGCGCCAACG	TGTCGGCGGC	GCGCGCATAC	TCGCCGTCTT
	1101	CGAGCCGCGT	TCCAACACCA	TGAAACTCGG	CACGATGAAG	TCCGCGCTCG
	1151	CCGCAAGCCT	CAAAGAAGCC	GACCAAGTGT	TCTGCTACGG	CGCGCGCGCG
	1201	GACTGGGACG	TTGCCGAAGC	CCTCGCGCCT	TTGGGCTGCA	GGCTGCGCGT
	1251	CGGTAAAGAT	TTCGATACCT	TCGTTGCCGA	AATTGTGAAA	AACGCCCGAA
	1301	CCGGCGGACCA	TATTTTGGTG	ATGAGCAACG	GCGGTTTCGG	CGGAATACAC
	1351	ACCAACACTGC	TGGACGCTTT	GAGATAG		

20 This corresponds to the amino acid sequence <SEQ ID 874; ORF132ng-1>:

	1	MKHIHIIGIG	GTFMGGIAAI	AKEAGFKVSG	CDAKMYPPMS	TQLEALGIGV
	51	HEGFDAAQLE	EFQADIYVIG	NVARRGMDVV	EAILNRGLPY	ISGPQWLAEV
	101	VLHHHWVLGV	AGTHGKTTTA	SMLAWVLEYA	GLAPGFLIGG	VPENFGVSAR
25	151	LPQTPRQDPN	SKSPFFVIEA	DEYDTAFFDK	RSKFVHYRPT	TAVLNLEFDD
	201	HADIFADLGA	IQTQFHHLVR	TVPSEGLIVC	NGQQQSLQDT	LDKGCWTPEV
	251	KFGTGHGWQI	GEVNADGSFD	VLLDGKKAGH	VAWDLMGGHN	RMNALAVIAA
	301	ARHAGVDVQT	ACEALGAFKN	VKRMEIKGT	ANGITYVDDE	AHHPTAIETT
	351	IQGLRQRVGG	ARILAVLEPR	SNTMKLGTMK	SALPASLKEA	DQVFCYAGGA
30	401	DWDVAEALAP	LGCRLRVGKD	FDTFVAEIVK	NARTGDHILV	MSNGGGGGIH
	451	TKLLDALR*				

ORF132ng-1 and ORF132-1 show 93.2% identity in 458 aa overlap:

35 orf132ng-1.pep MKHIHIIGIGGTFMGGIAAIAKEAGFKVSGCDAKMYPPMSTQLEALGIGVHEGFDAQAQLE  
 orf132-1 MKHIHIIGIGGTFMGGIAAIAKEAGFEVSGCDAKMYPPMSTQLEALGIDVYEGFDDAQLD

40 orf132ng-1.pep EFQADIYVIGNVARRGMDVVEAILNRGLPYISGPQWLAEENVLHHHWVLGVAGTHGKTTTA  
 orf132-1 EFKADVYVIGNVAKRGMDVVEAILNLGLPYISGPQWLSENVLHHHWVLGVAGTHGKTTTA

45 orf132ng-1.pep SMLAWVLEYAGLAPGFLIGGVPENFGVSARLPQTPRQDPNSKSPFFVIEADEYDTAFFDK  
 orf132-1 SMLAWVLEYAGLAPGFLIGGVPENFGVSARLPQTPRQDPNSQSPPFFVIEADEYDTAFFDK

50 orf132ng-1.pep RSKFVHYRPRTAVLNNLEFDHADIFADLGAIQTQFHHLVVRTVPSEGLIVCNGQQQSLQDT  
 orf132-1 RSKFVHYRPRTAVLNNLEFDHADIFADLGAIQTQFHYLVVRTVPSEGLIVCNGRQQSLQDT

55 orf132ng-1.pep LDKGCTWPVEKFGTGHGWQIGEVNADGSGFDVLLDGKKAGHVAWDLMGGHNRMNALAVIAA  
 orf132-1 LDKGCTWPVEKFGTEHGWQAGEANADGSGFDVLLDGKTAGRVKWDLMGRHNRMNALAVIAA

60 orf132ng-1.pep ARHAGVDVQTACEALGAFKNVKRMEIKGTANGITVYDDFAHHPTAIETTIQGLRQRVGG  
 orf132-1 ARHVGVDIQTACEALGAFKNVKRMEIKGTANGITVYDDFAHHPTAIETTIQGLRQRVGG

orf132ng-1.pep ARILAVLEPRSNMTMKLGTMKXSALPASLKEADQVFCYAGGADWDVAEALAPLGCRRLRVGKD  
 orf132-1 ARILAVLEPRSNMTMKLGTMKXSALPVSLKEADQVFCYAGGVWDVAEALAPLGGRNLNVGKD

orf132ng-1.pep FDTFVAEIVKNARTGDHILVMSNGGFGGIHTKLLDALRX  
 orf132-1 FDAFVAEIVKNAEVDHILVMSNGGFGGIHTGKLLDALRX

In addition, ORF132ng-1 is homologous to a hypothetical *E.coli* protein:

pir||S56459 hypothetical protein o457 - Escherichia coli >gi|537075 (U14003)  
 ORF\_o457 [Escherichia coli] >gi|1790680 (AE000494) hypothetical 48.5 kD protein  
 in fbp-pmba intergenic region [Escherichia coli] Length = 457  
 Score = 474 bits (1207), Expect = e-133  
 Identities = 249/439 (56%), Positives = 294/439 (66%), Gaps = 13/439 (2%)

Query: 22 KEAGFKVSGCDKMYPPMSTQLEALGIGVHEGFDAQLEEFQADIYVIGNVARRGMDVVE 81  
 ++ G +V+G DA +YPPMST LE GI + +G+DA+QLE Q D+ +IGN RG VE  
 Sbjct: 21 RQLGHEVTGSDANVYPPMSTLLEKQGIELIQGYDASQLEP-QPDLVIIGNAMTRGNPCVE 79

Query: 82 AILNRGLPYISGPQWLAENVLHHHVVGLVAGTHGKTTTASMLAWVLEYAGLAPGFLIGGV 141  
 A+L + +PY+SGPQWL + VL WVL VAGTHGKTTTA M W+LE G PGF+IGGV  
 Sbjct: 80 AVLEKNIPYMSGPQWLHDFVLRDRWVLAVAGTHGKTTTAGMATWILEQCGYKPGFVIGGV 139

Query: 142 PENFGVSARLPQTPRQDPNSKSPFFVIEADEYDTAFFDKRSKFVHYRPTAVLNNLEFDH 201  
 P NF VSA L +S FFFVIEADEYD AFFDKRSKFVHY PRT +LNNLEFDH  
 Sbjct: 140 PGNFEVSAHL-----GESDFFVIEADEYDCAFFDKRSKFVHYCPRTLILNNLEFDH 190

Query: 202 ADIFADLGAIQTQFHHLVTRVPSEGLIVCNGQQQSLQDITLDKGCWTPVEKFGTGHGWQIG 261  
 ADIF DL AIQ QFHHLVR VP +G I+ +L+ T+ GCW+ E G WQ  
 Sbjct: 191 ADIFDDLKAIQKQFHHLVRIVPGQGRIIWPENDINLKQTMAMGCWSEQELVGEQGHWQAK 250

Query: 262 EVNADGS-FDVLLDGKKAGHVAWDLMGHNRMNALAVIAAARHAGVDVQTACEALGAFKN 320  
 ++ D S ++VLLDG+K G V W L+G HN N L IAAARH GV A ALG+F N  
 Sbjct: 251 KLTTDASEWEVLLDGEKVGEVKSILVGEHNMHNGLMAIAAARHVGVPADAANALGSFIN 310

Query: 321 VKRMEIKGTANGITVYDDFAHHTAIETITQGLRQRVGG-ARILAVLEPRSNTMKLGTM 379  
 +RR+E++G ANG+TVYDDFAHHTAI T+ LR +VGG ARI+AVLEPRSNTMK+G  
 Sbjct: 311 ARRLRLRGEANGVTVYDDFAHHTAILATLAALRGKVGGTARIIVLEPRSNTMKMGIC 370

Query: 380 KSALPASLKEADQVF-CYAGGADWDVAEALAPLGCRLRVGKDFDTFVAEIVKNARTGDHI 438  
 K L SL AD+VF W VAE D DT +VK A+ GDHI  
 Sbjct: 371 KDDLAPSLGRADEVFLLQPAHIFWQVAEVAEACVQPAHWSGDVDTLADMVVKTAQPGDHI 430

Query: 439 LVMSNGGFGGIHTKLLDAL 457  
 LVMSNGGFGGIH KLLD L  
 Sbjct: 431 LVMSNGGFGGIHQKLLDGL 449

Based on this analysis, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF132-1 (26.4kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 20A shows the results of affinity purification of the His-fusion protein, and Figure 20B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for FACS analysis (Figure 20C) and ELISA (positive result). These experiments confirm that ORF132 is a surface-exposed protein, and that it is a useful immunogen.

### Example 103

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 875>

1 ..CCGGGCTATT ACGGCTCGGA TGACGAATTT AAGCGGGCAT TCGGAGAAAA  
 51 CTCGCCGACA TmCAAGAAAC ATTGCAACCG GAGCTGCGGG ATTTATGAAC  
 101 CCGTATTGAA AAAATACGGC AAAAAGCGCG CCAACAACCA TTCGGTCAGC  
 151 ATTAGTGCGG ACTTCGGCGA TTATTTTCATG CCGTTCGCCA GCTATTCGCG  
 201 CACACACCGT ATGCCCAACA TCCAAGAAAT GTATTTTTC CAAATCGGCG  
 251 ACTCCGGCGT TCACACCGCC TTAACCAG AGCGCGCAA CACTTGGCAA  
 301 TTTGGCTTCr ATACCTATAA AAAAGGATTG TTAACAAG ATGATACATT

5  
10  
15

```
351 AGGATTAAAA CTGGTCGGCT ACCGCAGCCG CATCGACAAC TACATCCACA
401 ACGTTTACGG GAAATGGTGG GATTTGAACG GGGATATTCC GAGCTGGGTC
451 AGCAGCACCG GGCTTGCCTA CACCATCCAA CATCGCrATT TCaWAGACAA
501 AGTGCATCAA nnnnnnnnnn nnnnnnnnnn nnnnTACGAT TATGGGCGTT
551 TTTTCACCAA CCTTCTTAC GCCTATCAAA AAAGCACGCA ACCGACCAAC
601 TTCAGCGATG CGAGCGAATC GCCCAACAAT GCGTCCAAAG AAGACCAACT
651 CAAACAAGGT TATGGGTTGA GCAGGGTTTC CGCCCTGCCG CGAGATTACG
701 GACGTTTGA AGTCGGTACG CGCTGGTTGG GCAACAAACT GACTTTGGGC
751 GGCGCGATGC GCTATTTCCG CAAGAGCATC CGCGCGACGG CTGAAGAACG
801 CTATATCGAC GGCACCAACG GGGGAAATAC CAGCAATTTC CGGCAACTGG
851 GCAAGCGTTC CATCAAACAA ACCGAAACTC TTGCCCGCCA GCCTTTGATT
901 TTWgATTTTa ACGCCGCTTA CGAGCCGAAG AAAAACCTTA TTTTCCGCGC
951 CGAAGTCAAA AATCTGTTCG ACAGGCGTTA TATCGATCCG CTCGATCGCG
1001 GCAATGATGC GGCAAC.GAG CGTTATTACA GCTCGTTTCA CCCGAAGAC
1051 AAGGACrrAG ACGTAACGTG TAATGCTGAT AAAACGTTGT GCaACGGCAA
1101 ATACGGCGGC ACAAGCAAAA GCGTATTGAC CAATTTTGCA CGCGGACGCA
1151 CCTTTTTgAT GACGATGAGC TACAAGTTT AA
```

This corresponds to the amino acid sequence <SEQ ID 876; ORF133>:

20  
25

```
1 . . PGGYGSDDDEF KRAFGENSPT XKKHCNRS CG IYEPVLKKYG KKRANNHSVS
51 ISADFGDYFM PFASYSRTHR MPNIQEMYFS QIGDSGVHTA LKPERANTWQ
101 FGFXTYKKGL LKQDDTLGLK LVGYRSRIDN YIHNVYKWW DLNGDIPSWV
151 SSTGLAYTIQ HRXFXDKVHQ XXXXXXXXXD YGRFFTNLSY AYQKSTQPTN
201 FSDASESPNN ASKEDQLKQG YGLSRVSALP RDYGRLEVGT RWLGNKLTGL
251 GAMRYFGKSI RATAERYID GTNGNTSNF RQLGKRSIKQ TETLARQPLI
301 XDFNAAYEPK KNLI FRAEVK NLFDRRYIDP LDAGNDAAAXE RYSSSFDPKD
351 KDXDVT CNAD KTL CNKYGG TSKSVLTNFA RGRFLMTMS YKF*
```

Further work revealed the further partial DNA sequence <SEQ ID 877>:

30  
35  
40  
45  
50  
55  
60  
65

```
1 GAGGCGCAGA TACAGGTTTT GGAAGATGTG CACGTCAAGG CGAAGCGCGT
51 ACCGAAAGAC AAAAAAGTGT TTACCGATGC GCGTGCCGTA TCGACCCGTC
101 AGGATATATT CAAATCCAGC GAAAACCTCG ACAACATCGT ACGCAGCATC
151 CCCGGTGCGT TTACACAGCA AGATAAAAGC TCGGGCATTG TGTCTTTGAA
201 TATTCGCGGC GACAGCGGGT TCGGGCGGGT CAATACGATG GTGGACGGCA
251 TCACGCAGAC CTTTTATTTCG ACTTCTACCG ATGCGGGCAG GGCAGGCGGT
301 TCATCTCAAT TCGGTGCATC TGTGCAGAGC AATTTTATTG CCGGACTGGA
351 TGTGCTCAAA GGCAGCTTCA GCGGCTCGGC AGGCATCAAC AGCCTTGCCG
401 GTTCGGCGAA TCTGCGGACT TTAGGCGTGG ATGACGTCGT TCAGGGCAAT
451 AATACCTACG GCCTGCTGCT AAAAGGTCTG ACCGGCACCA ATTCAACCAA
501 AGGTAATGCG ATGGCGCGCA TAGGTGCGCG CAAATGGCTG GAAAGCGGAG
551 CATCTGTCGG TGTGCTTTAC GGGCACAGCA GGCGCAGCGT GGCGCAAAAT
601 TACGCGTGG GCGGCGGCGG GCAGCACATC GGAAATTTTG GCGCGGAATA
651 TTTGGAACGG CGCAAGCAGC GATATTTTGT ACAAGAGGGT GCTTTGAAAT
701 TCAATTCGA CAGCGGAAAA TGGGAGCGGG ATTTACAAAG GCAACAGTGG
751 AAATACAAGC CGTATAAAAA TTACAACAAC CAAGAACTAC AaAAATACAT
801 CGAAGAGCAT GACAAAAGCT GCGGGGAAAA CCTg.CaCCG CAATACGACA
851 TTACCCCAT CGATCCGTCC AGCCTGAAGC AGCAGTCGGC AGGCAATCTG
901 TTTAAATTGG AATACGACGG CGTATTCAAT AAATACACGG CGCAATTTTCG
951 CGATTTAAAC ACCAAAATCG GCAGCCGCAA AATCATCAAC CGCAATTATC
1001 AGTTCAATTA CGGTTTGTCT TTGAACCCGT ATACCAACCT CAATCTGACC
1051 GCAGCCTACA ATTCGGGCAG GCAGAAATAT CCGAAAGGGT CGAAGTTTAC
1101 AGGCTGGGGG CTTTTAAAGG ATTTTGAAC CTACAACAAC GCGAAAATCC
1151 TCGACCTCAA CAACACCGCC ACCTTCCGGC TGCCCCGCGA AACCAGATTG
1201 CAAACCACTT TGGGCTTCAA TTATTTCCAC AACGAATACG GCAAAAACCG
1251 CTTTCTTGAA GAATTGGGGC TGTTTTTCGA CGGTCCGTAT CAGGACAACG
1301 GGCTTTATTC CTATTTGGGG CGGTTTAAGG GCGATAAAGG GCTGCTGCCC
1351 CAAAAATCAA CCATTGTCCA ACCGCGCGGC AGCCAATATT TCAACACGTT
1401 TACTTTCGAT GCCGCGCTCA AAAAGACAT TTACCGCTTA AACTACAGCA
1451 CCAATACCGT CGGCTACCGT TTCGCGCGGC AATATACGGG CTATTACGGC
1501 TCGGATGACG AATTTAAGCG GGCATTCCGA GAAAACTCGC CGACATACAA
1551 GAAACATTGC AACC GGAGCT GCGGGATTTA TGAACCCGTA TTGAAAAAAT
1601 ACGGCAAAAA GCGCGCCAAC AACCATTCCG TCAGCATTAG TGCGGACTTC
1651 GCGGATTATT TCATGCCGTT CGCCAGCTAT TCGCGCACAC ACCGTATTGC
1701 CAACATCCAA GAAATGTATT TTTCCCAAT CGGCGACTCC GCGGTTTACA
1751 CCGCCTTAAA ACCAGAGCGC GCAAACACTT GGCAATTTGG CTTCAATACC
1801 TATAAAAAAG GATTGTTAAA ACAAGATGAT ACATTAGGAT TAAAACGGT
1851 CCGCTACCGC AGCCGCATCG ACAACTACAT CCACAACGTT TACGGGAAAT
1901 GGTGGGATTT GAACGGGGAT ATTCCGAGCT GGGTCAGCAG CACCGGGCTT
```

5  
10  
15

```

1951 GCCTACACCA TCCAACATCG CAATTTCAAA GACAAAGTGC ACAAACACGG
2001 TTTTGAGTTG GAGCTGAATT ACGATTATGG GCGTTTTTTC ACCAACCTTT
2051 CTTACGCCTA TCAAAAAAGC ACGCAACCGA CCAACTTCAG CGATGCGAGC
2101 GAATCGCCCA ACAATGCGTC CAAAGAAGAC CAACTCAAAC AAGGTTATGG
2151 GTTGAGCAGG GTTTCGCCCC TGCCGCGAGA TTACGGACGT TTGGAAGTCG
2201 GTACGCGCTG GTTGGGCAAC AAACGACTT TGGGCGGCGC GATGCGCTAT
2251 TTCGCAAGA GCATCCGCGC GACGGCTGAA GAACGCTATA TCGACGGCAC
2301 CAACGGGGGA AATACCAGCA ATTTCCGGCA ACTGGGCAAG CGTTCATCA
2351 AACAAACCGA AACTCTTGCC CGCCAGCCTT TGATTTTTGA TTTTACGCC
2401 GCTTACGAGC CGAAGAAAAA CCTATTTTC CGCGCCGAAG TCAAAAATCT
2451 GTTCGACAGG CGTTATATCG ATCCGCTCGA TGCGGGCAAT GATGCGGCAA
2501 CGCAGCGTTA TTACAGCTCG TTCGACCCGA AAGACAAGGA CGAAGACGTA
2551 ACGTGAATG CTGATAAAAC GTTGTGCAAC GGCAAATACG GCGGCACAAG
2601 CAAAAGCGTA TTGACCAATT TTGCACGCGG ACGCACCTTT TTGATGACGA
2651 TGAGCTACAA GTTTTAA

```

This corresponds to the amino acid sequence <SEQ ID 878; ORF133-1>:

20  
25  
30

```

1  EAQIQVLEDV HVKAKRVPKD KKVFTDARAV STRQDIFKSS ENLDNIVRSI
51  PGAFTQQDKS SGIVSLNIRG DSGFGRVNTM VDGITQTFYS TSTDAGRAGG
101 SSQFGASVDS NFIAGLDVVK GSFSGSAGIN SLAGSANLRT LGVDDVVQGN
151 NTYGLLLKGL TGTNSTKGNA MAAIGARKWL ESGASVGVLY GHSRRSVAQN
201 YRVGGGGQHI GNFGAEYLER RKQRYFVQEG ALKFNSDSGK WERDLQRQQW
251 KYKPYKNYNN QELQKYIEEH DKSWRENLXP QYDITPIDPS SLKQQSAGNL
301 FKLEYDGVFN KYTAQFRDLN TKIGSRKIIN RNYQFNYGLS LNPYTNLNLN
351 AAYNSGRQKY PKGSKFTGWG LLKDFETYNN AKILDNLNTA TFRLPRETEL
401 QTTLGFNYFH NEYGKNRFFE ELGLFFDGPD QDNGLYSYLG RFGDKGLLP
451 QKSTIVQFAG SOYFNTFYFD AALKKDIYRL NYSTNTVGYR FGGEYTGYYG
501 SDDEFKRAFG ENSPTYKKHC NRSCGIYEPV LKKYGKKRAN NHSVVISADF
551 GDYFMPFASY SRTHRMPNIQ EMYFSQIGDS GVHTALKPER ANTWFQFENT
601 YKKGLLKQDD TLGLKLVGYR SRIDNYIHNV YGKWWDLNGD IPSWVSSTGL
651 AYTIQHRNFK DKVHKHGFEL ELNYDYGRFF TNLSYAYQKS TQPTNFSNAS
701 ESPNNASKED QLKQGYGLSR VSALPRDYGR LEVGTRWLGK KLTLLGAMRY
751 FGKSIRATAE ERYIDGTNGG NTSNFRQLGK RSIKQTETLA RQPLIFDFYA
801 AYEPKKNLIF RAEVKNLFDR RYIDPLDAGN DAATQRYSS FDPKDKDEDV
851 TCNADKTLN GKYGGTSKSV LTNEFARGRT LMTMSYKF*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with the probable TonB-dependent receptor HI121 of *H. influenzae* (accession number U32801)

ORF133 and HI121 show 57% aa identity in 363aa overlap:

40  
45  
50  
55  
60

```

Orf133: 31 IYEPVLKKYGKKRANNSVVISADFGDYFMPFASYSRTHRMPNIQEMYFSQIGDSGVHTA 90
          I EP+L K G K+A NHS ++SA+ DYFMPF +YSRTHRMPNIQEM+FSQ+ ++GV+TA
HI121: 563 INEPILHKSGHKKAFNHSATLSAELSDYFMPFFYSRTHRMPNIQEMFESQVSNAQVNTA 622

Orf133: 91 LKPERANTWQFGFXYKKGLLKQDDTLGLKLVGYRSRIDNYIHNVYGKWWDLNGDIPSWV 150
          LKPE+++T+Q GF TYKKGL QDD LG+KLVGYRS I NYIHNVYG WW +P+W
HI121: 623 LKPEQSDTYQLGFNTYKKGLFTQDDVLGVKLVGYRSFIKNIHNVYGWW--RDGMPTWA 680

Orf133: 151 SSTGLAYTIQHRXFXDKVHXXXXXXXXXXDYGRFFTNLSYAYQKSTQPTNFSDAESPNN 210
          S G YTI H+ + V YD GRFF N+SYAYQ++ QPTN++DAS PNN
HI121: 681 ESNGFKYTIHQNYKPIVKKSGVELEINYDMGRFFANVSAYQRTNQPTNYADASPRPNN 740

Orf133: 211 ASKEDQLKQGYGLSRVSALPRDYGRLEVGTRWLGKLTLLGAMRYFGKSIRATAEERYID 270
          AS+ED LKQGYGLSRVS LP+DYGRLE+GTRW KLTLG A RY+GKS RAT EE YI+
HI121: 741 ASQEDILKQGYGLSRVSMPLKDYGRLELGTWFDQKLTGLAARYYGKSKRATIEEYIN 800

Orf133: 271 GTNGGNTSNFRQLGKRSIKQTETLARQPLIXDFNAAYEPKKNLIFRAEVKNLFDRIYIDP 330
          G+ + R+ ++K+TE + +QP+I D + +YEP K+LI +AEV+NL D+RY+DP
HI121: 801 GSR-FKNTLRRENYAVKKTEDIKKQPIILDHVSYPEIKDLIIKAEVQNLLDKRYVDP 859

Orf133: 331 LDAGNDAAXERYSSFDPKDKDXDVTNADKTLNKGKGGTSKSVLTNEFARGRTFLMTMS 390
          LDAGNDAA +RYYSS + + C D + C GG+ K+VL NFARGRT+++++
HI121: 860 LDAGNDAASQRYSSSL-----NNSIECAQDSSAC----GGSDKTVLYNFARGRTYILSLN 910

```

Orf133: 391 YKF 393

YKF

HI121: 911 YKF 913

5 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF133 shows 90.8% identity over a 392aa overlap with an ORF (ORF133a) from strain A of *N. meningitidis*:

10 orf133.pep PGYYGSDDEFKRAFGENSPTXKKHCNRSCGI  
 orf133a FYFDAALKKDIYRLNYSTNTVGYRFGGXGTGYXSDDEFKRAFGENSPTXKKHCNQSCGI  
 450 460 470 480 490 500

15 orf133.pep YEPVLKKYGKKRANNHVSISADFGDYFMPFASYSRTHRMPNIQEMYFSQIGDSGVHTAL  
 orf133a YEPVLKKYGKKRANNHVSISADFGDYFMPFASYSRTHRMPNIQEMYFSQIGDSGVHTAL  
 510 520 530 540 550 560

20 orf133.pep KPERANTWQFGFXTYKKGLLKQDDTLGLKLVGYRSRIDNYIHNVYGKWWDLNGDIPSWVS  
 orf133a KPERANTWQFGFNTYKKGLLKQDDILGLKLVGYRSRIDXYIHNVYGKWWDLNGNIPSWVS  
 570 580 590 600 610 620

25 orf133.pep STGLAYTIQHRXFXDKVHQXXXXXXXXXYDGRFFTNLSYAYQKSTQPTNFSDAESPNNAL  
 orf133a STGLAYTIQHRNFKDKVHKHGFLELNLDYXRFFTNLSYAYQKSTQPTNFSDAESPNNAL  
 630 640 650 660 670 680

30 orf133.pep SKEDQLKQGYGLSRVSALPRDYGRLEVGTWLGKNTLGGAMRYFGKSIRATAEERYIDG  
 orf133a SKEDQLKQGYGLSRVSALPRDYGRLEVGTWLGKNTLGGAMRYFGKSIRATAEERYIDX  
 690 700 710 720 730 740

35 orf133.pep TNGGNTSNFRQLGKRSIKQTETLARQPLIXDFNAAYEPKKNLIFRAEVKNLFDRRYIDPL  
 orf133a TNGXXTSNFRQLGKRSIXQTETLARQPLIFDXAYAYEPKXLIFFRAEVKNLFDRRYIDPL  
 750 760 770 780 790 800

40 orf133.pep DAGNDAAXERYYSFDPKDKDVTCTNADKTLNKGKYGTSKSVLTNFARGRTFLMTMSY  
 orf133a DAGNDAATQRYYSFDPKDKDEEVTCTNDDNTLCNGKYGGTSKSVLTNFARGXTFLITMSY  
 810 820 830 840 850 860

45 orf133.pep KFX  
 orf133a KFX  
 870

## 55 A partial ORF133a nucleotide sequence &lt;SEQ ID 879&gt; is:

1 AAAGACAAAA AAGTGTTTAC CGATGCGCGT GCCGTATCGA CCCGTCAGGA  
 51 TATATTCAAA TCCANCGAAA ACCTCGACAA CATCGTACGC ANCATCCCCG  
 101 GTGCGTTTAC ACANCAANAT AAAAGCTCGG GCNTTGTGTC TTTGAATATT  
 151 CGCNGCGACA GCGGGTTCGG GCGGGTCAAT ACNATGGTNG ACGGCATCAC  
 201 NCANACCTTT TATTCGACTT CTACCGATGC GGGCAGGGCA GCGGGTTTAT  
 251 CTCAATTCGG TGCATCTGTC GACAGCAATT TTATNGCCGG ACTGGATGTC  
 301 GTCAAAGGCA GCTTCAGCGG CTCGGCAGGC ATCAACAGCC TTGCCGGTTC  
 351 GCGGAATCTG CGGACTTTAN GCGTGGATGA TGTCGTTTCA GGAATANTA  
 401 CNTACGGCCT GCTGCTAAAA GGTCTGACCG GCACCAATTC AACCAAAGGT

5  
10  
15  
20  
25  
30  
35  
40

451 AATGCGATGG CGGCGATAGG TGC GCGCAAA TGGCTGAAA GCGGAGCATC  
501 TGTCGGTGTG CTTTACGGGC ACAGCAGGCG CAGCGTGGCG CAAAATTACC  
551 CGGTGGGCGG CGGCGGGCAG CACATCGGAA ATTTTGGCGC GGAATATCTG  
601 GAACGACGCA AGCAACGATA TTTTGAGCAA GAAGGCGGGT TGAAATTCAA  
651 TTCCAACAGC GGAATATGGG AGCGGGATTT CCAAAAGTCG TACTGGAAAA  
701 CCAAGTGGTA TCAAAAATAC GATGCCCCCG AAGAACTGCA AAAATACATC  
751 GAAGGTCATG ATAAAAGCTG GCGGGAAAAC CTGGCGCCGC AATACGACAT  
801 CACCCCATC GATCCGTCCA GCCTGAAGCN GCAGTCGGCA GGCAACCTGT  
851 TTAAATTGGA ATACGACGGC GTATTCAATA AATACACGGC GCAATTTTCG  
901 GATTTAAAC CCAAAATCGG CAGCCGCAA ATCATCAACC GCAATTATCA  
951 ATTCAATTAC GGTGTGTCTT TGAACCCGTA TACCAACCTC AATCTGACCG  
1001 CAGCTACAA TTCGGGCGAG CAGAAATATC CGAAAGGGTC GAAGTTTACA  
1051 GGTGGGGGGC TTTTNAAGA TTTTGAAACC TACAACAACG CAAAATCCT  
1101 CGACCTCANC AACACCTCCA CCTTCGGCT GCCCGTGAA ACCGAGTTGC  
1151 AAACCACTTT GGGCTTCAAT TATTTCCACA ACGAATACGG CAAAACCGC  
1201 TTTCTGAAG AATTGGGGCT GTTTTTCGAC GGTCCGGATC ANGACAACGG  
1251 GCTTTATTCC TATTTGGGGC GGTTAAGGG CGATAAAGGG CTGCTGCCCC  
1301 AAAAAATCAAC CATTGTCCAA CCGGCCGGA GCCAATATT CAACACGTT  
1351 TACTTCGATG CCGCGCTCAA AAAAGACATT TACCGCTTAA ACTACAGCAC  
1401 CAATACCGTC GGCTACCGTT TCGGCGGCNA ATATACGGGC TATTACNGCT  
1451 CGGATGACGA ATTTAAGCGG GCATTTCGAG AAAACTCGCC GACATACANG  
1501 AAACATTGCA ACCAGAGCTG CGGAATTTAT GAACCCGTAT TGAAAAATA  
1551 CGGCAAAAAG CGCGCCAACA ACCATTCCGT CAGCATTAGT GCGGACTTCG  
1601 GCGATTATTT CATGCCGTTT GCCAGCTATT CGCGCACACA CCGTATGCCC  
1651 AACATCCAAG AAATGTATTT TTCCCAAATC GCGGACTCCG GCGTTACAC  
1701 CGCCTTAAAA CCAGAGCGCG CAAACACTTG GCAATTTGGC TTCAATACCT  
1751 ATAAAAAGG ATTGTTAAAA CAAGATGATA TATTAGGATT AAAACTGGTC  
1801 GGTACCGCA GCGGCATCGA CNACTACATC CACAACGTTT ACGGGAAATG  
1851 GTGGGATTG AACGGGAATA TTCCGAGCTG GGTACGAGC ACCGGGCTTG  
1901 CCTACACCAT CCAACACCGC AATTTCAAAG ACAAAGTGCA CAAACACGGT  
1951 TTTGAGTTGG AGCTGAATTA CGATTATNGG CGTTTTTTCA CCAACCTTTC  
2001 TTACGCCTAT CAAAAAGCA CGCAACCGAC CAACTTCAGC GATGCGAGCG  
2051 AATCGCCCAA CAATGCGTCC AAAGAAGACC AACTCAAACA AGGTTATGGG  
2101 TTGAGCAGGG TTTCCGCCCT GCCGCGAGAT TACGGACGTT TGGAAGTCGG  
2151 TACGCGCTGG TTGGGCAACA AACTGACTTT GGGCGGCGCG ATGCGCTATT  
2201 TCGGCAAGAG CATCCGCGCG ACGGCTGAAG AACGCTATAT CGACGNCACC  
2251 AATCGGGNAN NATCCAGCAA TTTCCGGCAA CTGGGCAAGC GTTCCATCAN  
2301 ACAAACCGAA ACCCTTGCCC GCCAGCCTTT GATTTTTGAT TTNTACGCCG  
2351 CTTACGAGCC GAAGAAAAAN CTATTTTTC GCGCCGAAGT CAAAATCTG  
2401 TTCGACAGGC GTTATATCGA TCCGCTCGAT GCGGGCAATG ATGCGGCAAC  
2451 GCAGCGTTAT ACAGTTCGT TCGACCCGAA AGACAAGGAC GAAGAAGTAA  
2501 CGTGAATGA TGATAACAG TTATGCAACG GCAAATACGG CGGCACAAGC  
2551 AAAAGCGTAT TGACCAATTT TGCACGCGGA CNCACCTTTT TGATAACGAT  
2601 GAGCTACAAG TTTTAA

45 This encodes a protein having (partial) amino acid sequence <SEQ ID 880>:

50  
55  
60

1 KDKKVFTDAR AVSTRQDIFK SXENLDNIVR XIPGAFTXQX KSSGXVSLNI  
51 RXDSGFGRVN TMVDGITXTF YSTSTDAGRA GGSSQFGASV DSNFXAGLDV  
101 VKGSFSGSAG INSLAGSANL RTLXVDDVQ GNXYGLLLK GLTGTNSTKG  
151 NMAAIGARK WLESASVGV LYHSRRSVA QNYRVGGGGQ HIGNFGAEYL  
201 ERRKQRYFEQ EGGLKFNSNS GKWERDFQKS YWKTWKYQKY DAPQELQKYI  
251 EGHDKSWREN LAPQYDITPI DPSSLKXQSA GNLFKLEYDG VFNKYTAQFR  
301 DLNTKIGSRK IINRNYQFNY GLSLNPYTNL NLTAAYNSGR QKYPKGSKFT  
351 GWGLXKDFET YNNAKILDLX NTSTFRLPRE TELQTTLGFN YFHNEYGKNR  
401 FPEELGLFFD GPDNDGLYS YLGRFKGDKG LLPQKSTIVQ PAGSQYFNTF  
451 YFDAALKKDI YRLNYSTNTV GYRFGGXYTG YYXSDDEFKR AFGENSPTYX  
501 KHCNQSCGIY EPVLKKYGGK RANNHSVSIS ADFGDYFMPF ASYSRTHRMP  
551 NIQEMYFSQI GDSGVHTALK PERANTWQFG FNTYKKGLLK QDDILGLKLK  
601 GYRSRIDXYI HNRYGKWWDL NGNIPSWVSS TGLAYTIQHR NFKDKVHKHG  
651 FELELNYDYX RFFTNLSYAY QKSTOPTNFS DASESPNNAS KEDQLKQGYG  
701 LSRVSALPRD YGRLEVGTWR LGNKLTLGGA MRYFGKSIRA TAEERYIDXT  
751 NGXTSNFRQ LGKRSIXQTE TLRQPLIFD XYAAEYEPKKX LIFRAEVKNL  
801 FDRYIDPLD AGNDAATQRY YSSFDPKDKD EEVTCNDNDT LCNGKYGGTS  
851 KSVLTNFARG XTFLITMSYK F\*

ORF133a and ORF133-1 show 94.3% identity in 871 aa overlap:

65

orf133a.pep

10 20 30 40  
KDKKVFTDARAVSTRQDIFKSXENLDNIVRXIPGAFTXQXKS



	orfl133-1	EAQIQVLEDVHVKA	KRV	PKDKKVFTDARAVSTRQDIFKSS	ENLDNIVRSIPGAFTQQDKS	
		10	20	30	40	50 60
5	orfl133a.pep	50	60	70	80	90 100
	orfl133-1	SGXVSLNIRXDSG	FGRVNTMVDGITXTFYSTSTDAGRAGGSSQFGASVDSNFXAGLDVVK			
10		70	80	90	100	110 120
	orfl133a.pep	110	120	130	140	150 160
	orfl133-1	GSFSGSAGINSLAGSANLRTLXVDDVVQGNXTYGLLLKGLTGNTSTKGNAMAAIGARKWL				
15		130	140	150	160	170 180
	orfl133a.pep	170	180	190	200	210 220
	orfl133-1	ESGASVGVLYGHSRRSVAQNYRVGGGQHHIGNFGAEYLERRKQRYFEQEGGLKFNSNSGK				
20		190	200	210	220	230 240
	orfl133a.pep	230	240	250	260	270 280
	orfl133-1	WERDFQKSYWKTQKYQKYDAPQELQKYLEGHDKSWRENLAPOYDITPIDPSSSLKXQSAGN				
25		250	260	270	280	290
	orfl133a.pep	290	300	310	320	330 340
	orfl133-1	LFKLEYDGVFNKYTAQFRDLNLTIGSRKIINRNYQFNGLSLNPTNLNLTAAYNSGRQK				
30		300	310	320	330	340 350
	orfl133a.pep	350	360	370	380	390 400
	orfl133-1	YPKGSKFTGWGLKDFETYNNAKILDNLNTATFRLPRETELQTTGLGFNYFHNEYGKNRFP				
35		360	370	380	390	400 410
	orfl133a.pep	410	420	430	440	450 460
	orfl133-1	EELGLFFDGPDXDNGLYSLGRFKGDKGLLPQKSTIVQAGSQYFNTFYFDAALKKDIYR				
40		420	430	440	450	460 470
	orfl133a.pep	470	480	490	500	510 520
	orfl133-1	LNYSTNTVGYRFGGXYTGYYXSDDEFKRAFGENSPTYXKHCNQSCGIYEPVLKKGKRA				
45		480	490	500	510	520 530
	orfl133a.pep	530	540	550	560	570 580
	orfl133-1	NNHSVSISADFGDYFMPFASYSRTHRMPNIQEMYFSQIGDSGVHTALKPERANTWQFGFN				
50		540	550	560	570	580 590
	orfl133a.pep	590	600	610	620	630 640
	orfl133-1	TYKKGLLKQDDTLGLKLVGYRSRIDXYIHNHYGKWWDLNNGNIPSWVSSTGLAYTIQHRNF				
55		600	610	620	630	640 650
	orfl133a.pep	650	660	670	680	690 700
	orfl133-1	KDKVHKHGFELNLDYGRFFTNLSYAYQKSTQPTNFSDAESPNNASKEDQLKQGYGLS				
60		660	670	680	690	700 710
	orfl133a.pep	710	720	730	740	750 760
	orfl133-1	KDKVHKHGFELNLDYGRFFTNLSYAYQKSTQPTNFSDAESPNNASKEDQLKQGYGLS				
65		710	720	730	740	750 760
	orfl133a.pep	760	770	780	790	800 810
	orfl133-1	KDKVHKHGFELNLDYGRFFTNLSYAYQKSTQPTNFSDAESPNNASKEDQLKQGYGLS				
70		760	770	780	790	800 810

-500-

		710	720	730	740	750	760
5	orf133a.pep	RVSALPRDYGRLEVGT	RWLGNKLT	LGAMRYFGKS	SIRATAEERY	IDXTNGXXTSNFRQLG	
	orf133-1	RVSALPRDYGRLEVGT	RWLGNKLT	LGAMRYFGKS	SIRATAEERY	IDGTNGGNTSNFRQLG	
		720	730	740	750	760	770
		770	780	790	800	810	820
10	orf133a.pep	KRSIXQTETLARQPLIF	DXYAAYEPKKLI	FRAEVKNLF	DRRYIDPLDAGNDAATQRYYS		
	orf133-1	KRSIXQTETLARQPLIF	DFYAAYEPKKLI	FRAEVKNLF	DRRYIDPLDAGNDAATQRYYS		
		780	790	800	810	820	830
		830	840	850	860	870	
15	orf133a.pep	SFDPKDKDEEVT	CNDNTLCNGKYGGT	SKSVLTNFARGXTFLITMSYKFX			
	orf133-1	SFDPKDKDEDVT	CNADKTL	CNGKYGGT	SKSVLTNFARGRTFLMTMSYKFX		
		840	850	860	870	880	

### Homology with a predicted ORF from *N.gonorrhoeae*

20 ORF133 shows 92.3% identity over 392 aa overlap with a predicted ORF (ORF133ng) from *N. gonorrhoeae*:

	orf133.pep	PGYYGSDDEFKRAFGENSPTXKKHCNRSCGI	31
	orf133ng	FYFDAALKKDIYRLNYSTNAINYRFGGEYTGYYGSENEFKRAFGENSPAYKEHCDPSCGL	560
25	orf133.pep	YEPVLKKYGKKRANNHVSISADFGDYFMPFASYSRTHRMPNIQEMYFSQIGDSGVHTAL	91
	orf133ng	YEPVLKKYGKKRANNHVSISADFGDYFMPFAGYSRTHRMPNIQEMYFSQIGDSGVHTAL	620
30	orf133.pep	KPERANTWQFGFXYTKGLLKQDDTLGLKLVGYRSRIDNYIHNVYGKWWDLNGDIPSWVS	151
	orf133ng	KPERANTWQFGFNTYKGLLKQDDILGLKLVGYRSRIDNYIHNVYGKWWDLNGDIPSWVG	680
35	orf133.pep	STGLAYTIQHRXFXDKVHQXXXXXXXXXYDYGRFFTNLSYAYQKSTQPTNFSDAESPNA	211
	orf133ng	STGLAYTIRHRNFKDKVHKHGFELNLDYGRFFTNLSYAYQKSTQPTNFSDAESPNA	740
40	orf133.pep	SKEDQLKQGYGLSRVSALPRDYGRLEVGT	271
	orf133ng	SKEDQLKQGYGLSRVSALPRDYGRLEVGT	800
45	orf133.pep	TNGGNTSNFRQLGKRSIKQTETLARQPLIXDFNAAYEPKKNLIFRAEVKNLFDRRYIDPL	331
	orf133ng	TNGGNTSNVRQLGKRSIKQTETLARQPLIFDYAAYEPKKNLIFRAEVKNLFDRRYIDPL	860
50	orf133.pep	DAGNDAAXERYYSFDPKDKDXDVT	391
	orf133ng	DAGNDAATQRYYSFDPKDKDEDVT	920
	orf133.pep	KF 393	
	orf133ng	KF 922	

The complete length ORF133ng nucleotide sequence <SEQ ID 881> is predicted to encode a protein having amino acid sequence <SEQ ID 882>:

55	1	MRSSFRLKPI	CFYLMGVMLY	HHSYAEDAGR	AGSEAQIQVL	EDVHVKA	KAKRV
	51	PKDKKVFTDA	RAVSTRQDVF	KSGENLDNIV	RSIPGAFTQQ	DKSSGIV	SLN
	101	IRGDSGFGRV	NTMVDGITQT	FYSTSTDAGR	AGGSSQFGAS	VDSNFIAG	L
	151	VVKGSFSGSA	GINSLAGSAN	LRTLGVDDV	QGNNTYGLLL	KGLTGTN	STK
	201	GNAMAAIGAR	KWLESGASVG	VLYGHSRRGV	AQNYRVGGGG	QHIGNFG	E
60	251	LERRKQQYFV	QEGGLKFNAG	SGKWERDLQR	QYWKTKWYKK	YEDPQEL	QKY
	301	IEEHDKSWRE	NLAPQYDITP	IDPSGLKQQS	AGNLLNLEYD	GVFNKYTA	Q
	351	RDLNTRIGSR	KIINRNYQFN	YGLSLNPYTN	LNLTAAYNSG	RQKYPKG	AKF
	401	TGWGLLKDFE	TYNNAKILDL	NNTATFRLPR	ETELQTTLGF	NYFHNEYG	KN

5  
10

451	RFPEELGLFF	DGPDQDNGLY	SYLGRFKGDK	GLLPQKSTIV	QPAGSQYFNT
501	FYFDAALKKD	IYRLNYSTNA	INRYFGGEYT	GYYGSENEFK	RAFGENSPAY
551	KEHCDPSCGL	YEPVLKKYK	KRANNHSVSI	SADFGDYFMP	FAGYSRTHRM
601	PNIQEMYFSQ	IGDSGVHTAL	KPERANTWQF	GFNTYKKGLL	KQDDILGLKL
651	VGYRSRIDNY	IHNVYGKWD	LNGDIPSWVG	STGLAYTIRH	RNEFKDKVHKH
701	GFELELNLDY	GRFFTNLSYA	YQKSTQPTNF	SDASESPNNA	SKEDQLKQGY
751	GLSRVSALPR	DYGRLEVGT	WLGKLTLLGG	AMRYFGKSIR	ATAEERYIDG
801	TNGGNTSNVR	QLGKRSIKQT	ETLARQPLIF	DFYAAEYFK	NLIFRAEVKN
851	LFDRRYIDPL	DAGNDAATQR	YYSSFDPKDK	DEDVTCNADK	TLCNGKYGGT
901	SKSVLTNFA	RGTFLMTMSY	KE*		

A variant was also identified, being encoded by the gonococcal DNA sequence <SEQ ID 883>:

15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65

1	ATGAGATCTT	CTTTCCGGTT	GAAGCCGATT	TGTTTTTATC	TTATGGGTGT
51	TATGCTATAT	CATCATAGTT	ATGCCGAAGA	TGCAGGGCGC	GCGGGCAGCG
101	AGGCGCAGAT	ACAGGTTTTG	GAAGATGTGC	ACGTCAGGCG	GAAGCGCGTA
151	CCGAAAGACA	AAAAAGTGTT	TACCGATGCG	CGTGCCGTAT	CGACCCGTca
201	gGATGTGTTT	AAATCCGGCG	AAAACCTCGA	CAACATCGTA	CGCAGCATAC
251	CCGGTGCGTT	TACACAGCAA	GATAAAAGCT	CGGGCATTGT	GTCTTTGAAT
301	ATTCGCGGCG	ACAGCGGGTT	CGGGCGGGTC	AATACGATGG	TGGACGCGAT
351	CACGCAGACC	TTTTATTCTA	CTTCTACCGA	TGCGGGCAGG	GCAGGCGGTT
401	CATCTCAATT	CGGTGCATCT	GTCGACAGCA	ATTTTATTGC	CGGACTGGAT
451	GTCGTCAAAG	GCAGCTTCAG	CGGCTCGGCA	GGCATCAACA	GCCTTGCCGG
501	TTGCGCGAAT	CTGCGGACTT	TAGGCGTGGA	TGACGTCGTT	CAGGGCAATA
551	ATACCTACGG	CCTGCTGCTA	AAAGGTCTGA	CCGGCACCAA	TTCAACCAA
601	GGTAATGCGA	TGGCGCGCAT	AGGTGCGCGC	AAATGGCTGG	AAAGCGGAGC
651	GTCTGTGCGT	GTGCTTTACG	GGCACAGCAG	GCGCGGCGTG	GCGCAAAATT
701	ACCGCGTGGG	CGGCGGCGGG	CAGCACATCG	GAAATTTTGG	TGAAGAATAT
751	CTGGAACGGC	GCAAAACAGCA	ATATTTTGTA	CAAGAGGGTG	GTTTGAATTT
801	CAATGCCGGC	AGCGGAAAAT	GGGAACGGGA	TTTGCAAAGG	CAATACTGGA
851	AAACAAAGTG	GTATAAAAAA	TACGAAGACC	CCCAAGAACT	GCAAAAATAC
901	ATCGAAGAGC	ATGATAAAAG	CTGGCGGGAA	AACCTGGGCG	CGCAATACGA
951	CATCACCCCC	ATCGATCCGT	CCGGCCTGAA	GCAGCAGTCG	GCAGGCAATC
1001	TGTTTAAATT	GGAATACGAC	GGCGTATTCA	ATAAATACAC	GGCGCAATTT
1051	CGCGATTTAA	ACACCAAGAT	CGGCAGCCGC	AAAATCATCA	ACCGCAATTA
1101	TCAATTCAT	TACGGTTTGT	CTTTGAACCC	GTATACCAAC	CTCAATCTGA
1151	CCGGAACCTA	CAATTCGGGC	AGGCAGAAAT	ATCCGAAAGG	GGCGAATTTT
1201	ACAGGCTGGG	GGCTTTTAAA	AGATTTTGAA	ACCTACAACA	ACGCGAAAAT
1251	CCTCGACCTC	AACAACACCG	CCACCTTCCG	GCTGCCCCGC	GAAACCGAGT
1301	TGCAAAACCAG	TTTGGGCTTC	AATTATTTCC	ACAACGAATA	CGGCAAAAAC
1351	CGCTTTCTCT	AAGAATTGGG	GCTGTTTTTC	GACGGTCTCT	ATCAGGACAA
1401	CGGGCTTTAT	TCCTATTTGG	GGCGGTTTAA	GGGCGATAAA	GGGCTGTTGC
1451	CTCAAAAATC	AACCATTTGT	CAACCGGCGC	GCAGCCAATA	TTTCAACACG
1501	TTCTACTTCG	ATGCGCGCT	CAAAAAAGAC	ATTTACCGCT	TAAACTACAG
1551	CACCAATGCA	ATCAACTACC	GTTTCGGCGG	CGAATATACG	GGCTATTACG
1601	GCTCGGAAAA	CGAATTAAAG	CGGGCATTCT	GAGAAAACCT	GCCGGCATAC
1651	AAGGAACATT	GCGACCCGAG	CTGCGGGCTT	TATGAACCCG	TATTGAAAAA
1701	ATACGGCAAA	AAGCGCGCCA	ACAACCATTC	GGTCAGCATT	AGTGCGGACT
1751	TCGGCGATTA	TTTCATGCCG	TTGCGCGGCT	ATTCGCGCAC	ACACCGTATG
1801	CCCAACATCC	AAGAAATGTA	TTTTTCCCAA	ATCGGCGACT	CCGGCGTTCA
1851	CACCGCCTTA	AAACCAGAGC	GCGCAAACAC	TTGGCAATTT	GGCTTCAATA
1901	CCTATAAAAA	AGGATTGTTA	AAACAAGATG	ATATATTAGG	ATTGAACTG
1951	GTCGGCTACC	GCAGCCGCAT	TGACAACCTAC	ATCCACAACG	TTTACGGGAA
2001	ATGGTGGGAT	TTGAACGGGG	ATATTCCGAG	CTGGGTGCGC	AGCACCGGGC
2051	TTGCCTACAC	CATCCGACAC	CGCAATTTCA	AAGACAAAGT	GCACAAACAC
2101	GGTTTTGAGC	TGGAGCTGAA	TTACGATTAT	GGGCGTTTTT	TCACCAACCT
2151	TTCTTACGCC	TATCAAAAAA	GCACGCAACC	GACCAATTTT	AGCGATGCGA
2201	GCGAATCGCC	CAACAATGCC	tccaaAGAAG	ACCAACTCAA	ACAAGGTTAT
2251	GGGCTGAGCA	GGGTTTCCGC	CCTGCCGCGA	GATTACGGAC	GTTTGGAGT
2301	CGGTACGCGC	TGGTTGGGCA	ACAAACTGAC	TTTGGGCGGC	GCGATgcGCT
2351	ATTTCCGGCAA	GAGCATCCGC	GCGACGGCTG	AAGAACGCTA	TATCGACGGC
2401	ACCAACGGGG	GAAATACCAG	CAATGTCCGG	CAACTGGGCA	AGCGTTCCAT
2451	AAAACAAACC	GAAACCCTTG	CCCGACAGCC	TTTGATTTTT	GATTTTTACG
2501	CCGCTTACGA	GCCGAAGAAA	AACCTTATTT	TCCGCGCCGA	AGTCAAAAAC
2551	CTGTTCGACA	GGCGTTATAT	CGATCCGCTC	GATGCGGGCA	ATGATGCGGC
2601	AACGCAGCGT	TATTACAGCT	CGTTCGACCC	GAAAGACAAG	GACGAAGACG
2651	TACCGTGTA	TGCTGATAAA	ACGTTGTGCA	ACGGCAAAATA	CGGCGGCACA
2701	AGCAAAAGCG	TATTGACCAA	TTTCGCACGC	GGACGCACCT	TCTTGATGAC
2751	GATGAGCTAC	AAGTTTTTAA			

This corresponds to the amino acid sequence <SEQ ID 884; ORF133ng-1>:

5  
10  
15  
20

```
1 MRSSFRLKPI CFYLMGVMLY HHSYAEDAGR AGSEAQIQVL EDVHVAKAKRV
51 PKDKKVFTDA RAVSTRQDVF KSGENLDNIV RSIPGAFTQQ DKSSSGIVSLN
101 IRGDSGFGRV NTMVDGITQT FYSTSTDAGR AGSSSQFGAS VDSNFIAGLD
151 VVKGSFSGSA GINSLAGSAN LRTLGVDDVV QGNNTYGLLL KGLTGTNSTK
201 GNAMAAIGAR KWLESGASVG VLYGHSRRGV AQNYRVGGGG QHIGNFGEY
251 LERRKQYFV QEGGLKFAG SGKWERDLQR QYWKTKWYK YEDPQELQKY
301 IEEHDKSWRE NLAPQYDITP IDPSGLKQOS AGNLFKLEYD GVFNKYTAQF
351 RDLNTRIGSR KIINRNYQFN YGLSLNPYTN LNLTAAYNSG RQKYPKGAKF
401 TGWGLLKDFE TYNNAKILD LNTATFRLPR ETELQTTLGF NYFHNEYGKN
451 RFPEELGLFF DGPQDNGLY SYLGRFKGDK GLLPQKSTIV QPAGSQYFNT
501 FYFDAALKKD IYRLNYSTNA INYRFGGEYT GYSGSENEFK RAFGENSPAY
551 KEHCDPSCGL YEPVLKKGK KRANNHVSIS SADFGDYFMP FAGYSRTHRM
601 PNIQEMYFSQ IGDGVHTAL KPERANTWQF GFNTYKKGLL KQDDILGLKL
651 VGYRSRIDNY IHNVYKQWWD LNDIPSWVG STGLAYTIRH RNFKDKVHHK
701 GFELELNDYD GRFFTNLSYA YQKSTQPTNF SDASESPNNA SKEDQLKQGY
751 GLSRVSALPR DYGRLEVGR WLGNKLTLLG AMRYFGKSIR ATAEERYIDG
801 TNGGNTSNVR QLGKRSIKQT ETLARQPLIF DFYAAYEPKK NLIFRAEVKN
851 LFDRIYIDPL DAGNDAATQR YYSSFDPKDK DEDVTCNADK TLCNGKYGGT
901 SKSVLTNFA GRFTLMTMSY KF*
```

ORF133ng-1 and ORF133-1 show 96.2% identity in 889 aa overlap:

600 500 400 300 200 100 0

```
10 20 30 40 50 60
orf133ng-1.pep SFRLLKPICFYLMGVMLYHHSYAEDAGRAGSEAQIQVLE
25 orf133-1 EAQIQVLE
10 20 30
70 80 90 100 110 120
orf133ng-1.pep STRQDVFKSGENLDNIVRSIPGAFTQQDKSSGIVSLNIRGDSGFGRVNTMVDGITQTFYS
30 orf133-1 STRQDIFKSSENLDNIVRSIPGAFTQQDKSSGIVSLNIRGDSGFGRVNTMVDGITQTFYS
40 50 60 70 80 90
130 140 150 160 170 180
orf133ng-1.pep TSTDAGRAGSSSQFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTLGVDDVVQGN
40 orf133-1 TSTDAGRAGSSSQFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTLGVDDVVQGN
100 110 120 130 140 150
190 200 210 220 230 240
orf133ng-1.pep NTYGLLLKGLTGTNSTKGNAMAAIGARKWLESGASVGVLVYGHSSRRVQAQNYRVGGGGQHI
45 orf133-1 NTYGLLLKGLTGTNSTKGNAMAAIGARKWLESGASVGVLVYGHSSRRVQAQNYRVGGGGQHI
160 170 180 190 200 210
250 260 270 280 290 300
orf133ng-1.pep GNFGEEYLERRKQYFVQEGGLKFAGSGKWERDLQRQYWKTKWYKYYEDPQELQKYIEE
50 orf133-1 GNFGAEYLERRKQRYFVQEGALKFNSDSGKWERDLQRQYWKTKWYKYYEDPQELQKYIEE
220 230 240 250 260
310 320 330 340 350 360
orf133ng-1.pep HDKSWRENLAPOYDITPIDPSGLKQOSAGNLFKLEYDGVFNKYTAQFRDLNTRIGSRKII
55 orf133-1 HDKSWRENLAPOYDITPIDPSGLKQOSAGNLFKLEYDGVFNKYTAQFRDLNTRIGSRKII
270 280 290 300 310 320
370 380 390 400 410 420
orf133ng-1.pep NRNYQFNYGLSLNPYTNLNLTAAYNSGRQKYPKGAKFTGWGLLKDFETYNNAKILDNLNT
60 orf133-1 NRNYQFNYGLSLNPYTNLNLTAAYNSGRQKYPKGAKFTGWGLLKDFETYNNAKILDNLNT
330 340 350 360 370 380
430 440 450 460 470 480
orf133ng-1.pep ATFRLPRETELQTTLGFNYFHNEYGKNRFPEELGLFFDGPQDNGLYSYLGRFKGDKGLL
65 orf133-1 ATFRLPRETELQTTLGFNYFHNEYGKNRFPEELGLFFDGPQDNGLYSYLGRFKGDKGLL
```

[illegible]

In addition, ORF133ng-1 is homologous to a TonB-dependent receptor in *H.influenzae*:

sp|P45114|YC17\_HAEIN PROBABLE TONB-DEPENDENT RECEPTOR HI1217 PRECURSOR  
>gi|1075372|pi|G64110 transferrin binding protein 1 precursor (tbpl) homolog -  
55 Haemophilus influenzae (strain Rd KW20) >gi|1574147 (U32801) transferrin binding  
protein 1 precursor (tbpl) [Haemophilus influenzae] Length = 913  
Score = 930 bits (2377), Expect = 0.0  
Identities = 476/921 (51%), Positives = 619/921 (66%), Gaps = 72/921 (7%)

Query: 38 QVLEDVHVKAKRVPKDKKVFSTDARAVSTRQDVFKSGENLDNIVRSIPGAFTQQDKSSGIV 97  
+ L + V K + DKK FT+A+A STR++VFK + +D ++RSIPGAFTQQDK SG+V  
60 Sbjct: 29 ETLGQIDVVEKVISNDKKPFTEAKAKSTRENVFKETQTIDQVIRSIPGAFTQQDKSGSVV 88

Query: 98 SLNIRGDSGGRVNTMVDGITQTFYSTSTDAGRAGSSQFGASVDSNFIAGLDVVKGSFS 157  
+S+NIRG++G GRVNTMVDG+TQTFYST+ D+G++GGSSQFGA++D NFIAG+DV K +FS  
65 Sbjct: 89 SVNIRGENGLGRVNTMVDGVTQTFYSTALDSGQSGSSQFGAAIDPNFIAGVDVNKSNFS 148

Query: 158 GSAGINSLAGSANRLTLGVDDVVQXXXXXXXXXXXXXXXXXXXXXAMAAIGARKWLESGA 217  
G++GIN+LAGSAN RTLGV+DV+ M RKWJ++G

Sbjct: 149 GASGINALAGSANFRTLGVNDVITDDKPFGIILKGMTGSNATKSNFMTMAAGRKWLNDGG 208  
 Query: 218 SVGVLYGHSRRGVAQNYRVGGGQHIGNFGEEYLERRKQOYFVQEGGLKFNAGSGKWERD 277  
 VGV+YG+S+R V+Q+YR+ GGG+ + + G++ L + K+ YF + G N G+W D  
 Sbjct: 209 YGVVYGYSQREVSQDYRI-GGGERLASLQDILAKEKEAYF-RNAGYILNP-EGQWTPD 265  
 Query: 278 LQRQYWK-----TKWY-----KKYEDPQELQK---YIEE 303  
 L +++W +Y KK +D ++LQK IEE  
 Sbjct: 266 LSKKHWSCNKPDPYQKNGDCSYRIGSAAKTRREILQELLTNGKKPKDIEKLQKNGDIEE 325  
 Query: 304 HDKSWRENLAPOYDITPIDPSGLKQOSAGNLFKLEYDGVFNKYTAQFRDLNTRIGSRKII 363  
 DKS+ N QY + PI+P L+ +S +L K EY AQ R L+ +IGSRKI  
 Sbjct: 326 TDKSFERN-KDQYSVAPIEPGSLQSRSRSHLLKFEYGDHQNLAGQLRTLNDKIGSRKIE 384  
 Query: 364 NRNYQFNYGLSLNPYTNLNLTAAYNSGRQKYPKGAKFTGWLLKDFETYNNAKILDNLNT 423  
 NRNYQ NY + N Y +LNL AA+N G+ YPKG F GW + T N A I+D+NN+  
 Sbjct: 385 NRNYQVNYNFNNSYLDLNLMAAHNIGKTIYPKGGFFAGWQVADKLITKNVANIVDINNS 444  
 Query: 424 ATFRLPRETELQTTLGFNHYFNEYGKNRFPPEELGLFFDGPQDNGLYSY--LGRFKGDKG 481  
 TF LP+E +L+TTLGFNHYF NEY KNRFPPEEL LF++ D GLYS+ GR+ G K  
 Sbjct: 445 HTFLLPKEIDLKTTLGFNHYFTNEYSKNRFPPEELSLFYNDASHDQGLYSHSKRGYSGTKS 504  
 Query: 482 LLPQKSTIVQPAGSQYFNTFYFDAALKKDIYRLNYSTNAINYRFGGEYTGYYGSENEFKR 541  
 LLPQ+S I+Q+G Q F T YFD AL K IY LNYS N +Y F GEY GY  
 Sbjct: 505 LLPQRSVILQPSGKQKFKTVYFDTALSKGIYHLNYSVNFTHYAFNGEYVGY----- 555  
 Query: 542 AFGENSPAYKEHCDPSCGLYEPVLKKGKRRANNHVSISADFGDYFMPFAGYSRTHRMP 601  
 EN+ + + EP+L K G K+A NHS ++SA+ DYFMPF YSRTHRMP  
 Sbjct: 556 ---ENTAGQQ-----INEPILHKS GHKKA FNHSATLSAELSDYFMPFFTYSRTHRMP 604  
 Query: 602 NIQEMYFSQIGDSGVHTALKPERANTWQFGFNTYKKGLLKQDDILGLKLVGYRSRIDNYI 661  
 NIQEM+FSQ+ ++GV+TALKPE+++T+Q GFNTYKKGL QDD+LG+KLVGYRS I NYI  
 Sbjct: 605 NIQEMFQSQVSNAGVNTALKPEQSDTYQLGFNTYKKGLFTQDDVLGVKLVGYRSFKNYI 664  
 Query: 662 HNRYGKWWDLNGDIPSWVGSTGLAYTIRHRNFKDKVHKHGFELNLYDYGRFFTNLSYAY 721  
 HNRYG WW +P+W S G YTI H+N+K V K G ELE+NYD GRFF N+SYAY  
 Sbjct: 665 HNRYGVWW--RDGMPTWAESNGFKYTIHQNYKPIVKKSGVELEINYDMGRFFANVSYAY 722  
 Query: 722 QKSTQPTNFSDASESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGTWLGKLTGLGA 781  
 Q++ QPTN++DAS PNNAS+ED LKQGYGLSRVS LP+DYGRLE+GTRW KLTGL A  
 Sbjct: 723 QRTNQPTNYADASPRPNNASQEDILKQGYGLSRVSMPLPKDYGRLELGTWFDQKLTGLA 782  
 Query: 782 MRYFGKSIRATAEERYIDGTNGGNTSNVRQLGKRSIKQTETLARQPLIFDFYAAEYEPKKN 841  
 RY+GKS RAT EE YI+G+ + +R+ ++K+TE + +QP+I D + +YEP K+  
 Sbjct: 783 ARYYGKSKRATIEEEYINGSR-FKKNLRRNNYAVKKTEDIKKQPIILDHVSYEPIKD 841  
 Query: 842 LIFRAEVKNLFDRIYIDPLDAGNDAATQRYYSFDPKDKDEDVTCNADKTLNKGKYGTS 901  
 LI +AEV+NL D+RY+DPLDAGNDA+QRYYS + + C D + C GG+  
 Sbjct: 842 LIIKAEVQNLLDKRYVDPDAGNDAASQRYYSL-----NNSIECAQDSSAC----GGSD 892  
 Query: 902 KSVLTNFARGRTFLMTMSYKF 922  
 K+VL NFARGRT+++++YKF  
 Sbjct: 893 KTVLYNFARGRTYILSLNYKF 913

The underlined motif in the gonococcal protein (also present in the meningococcal protein) is  
 predicted to be an ATP/GTP-binding site motif A (P-loop), and the analysis suggests that these  
 proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for  
 vaccines or diagnostics, or for raising antibodies.

#### Example 104

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 885>

1 ATGAACCTGA TTTCACGTTA CATCATCCGT CAAATGGCGG TTATGGCGGT

51 TTACGCGCTC CTTGCCTTCC TCGCTTTGTA CAGCTTTTTT GAAATCCTGT  
 101 ACGAAACCGG CAACCTCGGC AAAGGCAGTT ACGGCATATG GGAAATGCTG  
 151 GGCTACACCG CCTCAAAAT GCCCGCCCGC GCCTACGAAC TGATTCCCCT  
 201 CGCCGTCTTT ATCGGCGGAC TGGTCTCCCT CAGCCAGCTT GCCGCCGGCA  
 251 GCGAACTGAC CGTCATCAA GCCAGCGGCA TGAGCACCAA AAAGCTGCTG  
 301 TTGATTCTGT CGCAGTTCGG TTTTATTTT GCTATTGCCA CCGTCGCGCT  
 351 CGGCGAATGG GTTGCGCCCA CACTGAGCCA AAAAGCCGAA AACATCAAAG  
 401 CCGCCGCCAT CAACGGCAAA ATCAGCACCG GCAATACCGG CCTTTGGCTG  
 451 AAAGAAAAA ACAGCGTGAT CAATGTGCGC GAAATGTTGC CCGACCAT..

10 This corresponds to the amino acid sequence <SEQ ID 886; ORF112>:

1 MNLISRYIIR QMAVMAVYAL LAFLALYSFF EILYETGNLG KGSYGIWEML  
 51 GYTALKMPAR AYELIPLAVL IGGLVLSLSQL AAGSELTVIK ASGMSTKKLL  
 101 LILSQFGFIF AIATVALGEW VAPTLSQKAE NIKAAAINGK ISTGNTGLWL  
 151 KEKNSVINVR EMLPDH...

15 Further work revealed further partial nucleotide sequence <SEQ ID 887>:

1 ATGAACCTGA TTTCACGTTA CATCATCCGT CAAATGGCGG TTATGGCGGT  
 51 TTACGCGCTC CTTGCCTTCC TCGCTTTGTA CAGCTTTTTT GAAATCCTGT  
 101 ACGAAACCGG CAACCTCGGC AAAGGCAGTT ACGGCATATG GGAAATGCTG  
 151 ggCTACACCG CCTCAAAAT GCCCGCCCGC GCCTACGAAC TGATTCCCCT  
 201 CGCCGTCTTT ATCGGCGGAC TGGTCTCCCT CAGCCAGCTT GCCGCCGGCA  
 251 GCGAACTGAC CGTCATCAA GCCAGCGGCA TGAGCACCAA AAAGCTGCTG  
 301 TTGATTCTGT CGCAGTTCGG TTTTATTTT GCTATTGCCA CCGTCGCGCT  
 351 CGGCGAATGG GTTGCGCCCA CACTGAGCCA AAAAGCCGAA AACATCAAAG  
 401 CCGCCGCCAT CAACGGCAAA ATCAGCACCG GCAATACCGG CCTTTGGCTG  
 451 AAAGAAAAA ACAGCTkAT CAATGTGCGC GAAATGTTGC CCGACCATAC  
 501 ACCTTTGGGC ATCAAAATTT GGGCGCGCAA CGATAAAAC GAATTGGCAG  
 551 AGGCAGTGGG AGCCGATTCC GCCGTTTGA ACAGCGACGG CAGTTGGCAG  
 601 TTGAAAAACA TCCGCCCGAG CACGCTTGGC GAAGACAAAG TCGAGGCTC  
 651 TATTGCGGCT GAAGAAACT GGCCGATTTC CGTCAAACGC AACCTGATGG  
 701 ACGTATTGCT CGTCAAACCC GACCAAATGT CCGTCGGCGA ACTGACCACC  
 751 TACATCCGCC ACCTCCAAA CAACAGCCAA AACACCCGAA TCTACGCCAT  
 801 CGCATGGTGG CGCAAATTGG TTTACCCCGC CGCAGCCTGG GTGATGGCGC  
 851 TCGTCGCCTT TGCCTTTACC CCGCAAACCA CCCGCCACGG CAATATGGGC  
 901 TTAAACTCT TCGGCGGCAT CTGTsTCGGA TTGCTGTTCC ACCTTGCCGG  
 951 ACGGCTCTTT GGGTTTACCA GCCAACTCGG...

This corresponds to the amino acid sequence <SEQ ID 888; ORF112-1>:

1 MNLISRYIIR QMAVMAVYAL LAFLALYSFF EILYETGNLG KGSYGIWEML  
 51 GYTALKMPAR AYELIPLAVL IGGLVLSLSQL AAGSELTVIK ASGMSTKKLL  
 101 LILSQFGFIF AIATVALGEW VAPTLSQKAE NIKAAAINGK ISTGNTGLWL  
 151 KEKNSXINVR EMLPDHTLLG IKIWARNDKN ELAEAVEADS AVLNSDGSWQ  
 201 LKNIRRLSTLG EDKVEVSIAA EENWPISVKR NLMDVLLVKP DQMSVGELTT  
 251 YIRHLQNNNSQ NTRIYAIAWW RKLVPFAAAW VMALVAFAFT PQTRRHNGM  
 301 LKLEGGICXG LLEHLAAGRLF GFTSQL...

Computer analysis of this amino acid sequence predicts two transmembrane domains and gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF112 shows 96.4% identity over a 166aa overlap with an ORF (ORF112a) from strain A of *N.meningitidis*:

		10	20	30	40	50	60
50	orf112.pep	MNLISRYIIRQMAVMAVYALLAFLALYSFFEILYETGNLGKGSYGIWEMLGYTALKMPAR					
	orf112a	MNLISRYIIRQMAVMAVYALLAFLALYSFFEILYETGNLGKGSYGIWEMLGYTALKMXAR					
		10	20	30	40	50	60
55	orf112.pep	AYELIPLAVLIGGLVLSLSQLAAGSELTVIKASGMSTKKLLLILSQFGFIFAIATVALGEW					
		70	80	90	100	110	120

		:
	orf112a	AYELMPLAVLIGLVXSXSQLAAGSELXVIKASGMSTKKLLLSQFGFIFAIATVALGEW
		70 80 90 100 110 120
5		130 140 150 160
	orf112.pep	VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKNVINVREMLPDH
	orf112a	VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKNSIINVREMLPDHTLLGIKIWARNDKN
		130 140 150 160 170 180
10		
	orf112a	ELAEAVEADSAVLNSDGSQWLKNIRRSTLGEDKVEVSIAAEEXWPISVKRNLM DVLLVKP
		190 200 210 220 230 240

The ORF112a nucleotide sequence <SEQ ID 889> is:

15	1	ATGAACCTGA	TTTCACGTTA	CATCATCCGT	CAAATGGCGG	TTATGGCGGT
	51	TTACGCGCTC	CTTGCCCTTC	TCGCTTTGTA	CAGCTTTTTT	GAAATCCTGT
	101	ACGAAACCGG	CAACCTCGGC	AAAGGCAGTT	ACGGCATAATG	GGAATGNTG
	151	GGNTACACCG	CCCTCAAAAT	GNCCGCCCGC	GCCTACGAAG	TGATGCCCT
20	201	CGCCGTCCTT	ATCGGCGGAC	TGGTCTCTNT	CAGCCAGCTT	CGCCGCCGCA
	251	GCGAACTGAN	CGTCATCAA	GCCAGCGGCA	TGAGCACCAA	AAAGCTGCTG
	301	TTGATTCTGT	CGCAGTTCCG	TTTATTTTTT	GCTATTGCCA	CCGTCCGCTG
	351	CGCGGAATGT	GTTGCGCCCA	CACTGAGCCA	AAAAGCCGAA	AACATCAAAG
25	401	CGCGGCCCAT	CAACGGCAAA	ATCAGTACCG	GCAATACCGG	CCCTTGGCTG
	451	AAAGAAAAAA	ACAGCATTAT	CAATGTGCGC	GAAATGTTGC	CCGACCATAC
	501	CCTGCTGGGC	ATTAATAATCT	GGCCCCGCAA	CGATAAAAAAC	GAACTGGCAG
	551	AGGCAGTGGA	AGCCGATTCC	GCCGTTTTGA	ACAGCGACGG	CAGTTGGCAG
30	601	TTGAAAAACA	TCCGCCGAG	CACGCTTGGC	GAAGACAAAG	TCGAGTCTCT
	651	TATTGCGGCT	GAAGAAAANT	GGCCGATTTC	CGTCAAACGC	AACCTGATGG
	701	ACGTATTGCT	CGTCAAACCC	GACCAAATGT	CCGTCCGCGA	ACTGACCACC
	751	TACATCCGCC	ACCTCCAAAN	NNACAGCCAA	AACACCCGAA	TCTACGCCAT
35	801	CGCATGGTGG	CGCAAAATTGG	TTTACCCCGC	CGCAGCCTGT	GTCGGCGCG
	851	TCGTGCGCTT	TGCCTTTTACC	CCGCAAACCA	CCCGCCACGG	CAATATGGGC
	901	TTAAAANTCT	TCGGCGGCAT	CTGTCTCGGA	TTGTGTGTCC	ACCTTGCCGG
	951	NCGGCTCTTC	NGGTTTACA	GCCAACTCTA	CGGCATCCCG	CCCTTCCTCG
	1001	NCGGCGCACT	ACCTACCATA	GCCTTCGCCT	TGCTCGCCGT	TTGGCTGATA
	1051	CGCAAAACAGG	AAAAACGCTA	A		

This encodes a protein having the amino acid sequence <SEQ ID 890>:

```

1  MNLISRYIIR QMAVMAYVAL LAFLALYSFF EILYETGNLG KSGYGIWEMX
51 GYTALKMXAR AYELMPLAVL IGGLEVSXSQL AAGSELXVIK ASGMSTFKLL
101 LILSQFGFIF AIATVALGEG VAPTLSQKAE NIKAAAINGK ISTGNTGLWL
151 KEKNSTINVR EMLPDHTLGW IKIWARNDNK ELAEVAEADS AVLNSDGSQV
201 LKNIRRSTLG EDKVEVSIAA EEXWPISVKR NLMVDVLLVKP DQMSVGELTT
251 YIRHLQXXSQ NTRIYAIAWW RKLVPYAAAW VMALVAFaft PQTTRHGNMG
301 LKXFGGICLG LLFHLAGRLF XFTSQLYGIP PFLXGALPTI AFALLAVWLI
351 RKOEKR*

```

45 ORF112a and ORF112-1 show 96.3% identity in 326 aa overlap:

	orf112a.pep	MNLISRYIIQMAVMVAVYALLAFLALYSFFEILYETGNLGKGSYGIWEMXGYTALKMXAR 
	orf112-1	MNLISRYIIQMAVMVAVYALLAFLALYSFFEILYETGNLGKGSYGIWEMLGYTALKMPAR 
50	orf112a.pep	AYELMPLAVLIGGLVXSXSQLAAGSELXVIKASGMSTKKLLLLILSQFGFIFAIATVALGEW    :
	orf112-1	AYELIPLAVLIGGLVSLSQLAAGSELTVIKASGMSTKKLLLLILSQFGFIFAIATVALGEW    :
55	orf112a.pep	VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKNSIINVREMLPDHTLLGIKIWARNDKN 
	orf112-1	VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKNSXINVREMLPDHTLLGIKIWARNDKN 
	orf112a.pep	ELAEAVEADSAVLNSDGSWQLKNIRRSTLGEDKVEVSIAAEEXWPISVKRNLMDVLLVKP 
60	orf112-1	ELAEAVEADSAVLNSDGSWQLKNIRRSTLGEDKVEVSIAAEENWPISVKRNLMDVLLVKP 
	orf112a.pep	DQMSVGELTTYIRHLQXXSQNTRIYAIAWWRKLVYPAAAWMALVAFATPQTTRHGNMG 



orf112-1 DQMSVGELTTYIRHLQNNNSQNTRIYAIWWRKLVYPAAAWVMALVAFATPQTTRHGNMG  
 orf112a.pep LKXFGGICLGLLFHLAGRLFXFTS QLYGIPFLXGALPTIAFALLAVWLIRKQEKRX  
 5 orf112-1 LKLFGGICXGLLFHLAGRLFGFTSQL

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF112 shows 95.8% identity over 166aa overlap with a predicted ORF (ORF112ng) from *N. gonorrhoeae*:

10 orf112.pep MNLISRYIIRQMAVMNAVYALLAFLALYSFFEILYETGNLGKGSYGIWEMLGYTALKMPAR 60  
 orf112ng MNLISRYIIRQMAVMNAVYALLAFLALYSFFEILYETGNLGKGSYGIWEMLGYTALKMPAR 60  
 15 orf112.pep AYELIPLAVLIGGLVLSLSQLAAGSELTVIKASGMSTKKLLILSQFGFIFAIATVALGEW 120  
 orf112ng AYELMPLAVLIGGLASLSQLAAGSELAVIKASGMSTKKLLILSQFGFIFAIAAVALGEW 120  
 orf112.pep VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKNSVINVREMLPDH 166  
 20 orf112ng VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKTSIINVRGMLPDHTLLGIKIWARNDKN 180

The complete length ORF112ng nucleotide sequence <SEQ ID 891> is:

1 ATGAACCTGA TTTCACGTTA CATCATCCGC CAAATGGCGG TTATGGCGGT  
 51 TTACGCGCTC CTTGCCTTCC TCGCTTTGTA CAGCTTTTTT GAAATCCTGT  
 101 ACGAAACCGG CAACCTCGGC AAAGGCAGTT ACGGCATATG GGAAATGCTG  
 151 GGCTACACCG CCTCAAAAT GCCCGCCCGC GCCTACGAAC TCATGCCCTC  
 201 CGCCGTCCTC ATCGGCGGAC TGGCCTCTCT CAGCCAGCTT GCCGCCGCGA  
 251 GCGAACTGGC CGTCATCAAA GCCAGCGGCA TGAGCACCAA AAAGCTGCTG  
 301 TTGATTCTGT CTCAGTTCGG TTTATTTTTT GCTATGCGCG CCGTCGCGCT  
 351 CGGCGAATGG GTTGCGCCCA CGCTGAGCCA AAAAGCCGAA AACATCAAG  
 401 cCGCGCCAt taacggCAAA ATCAGCAccg gCAATACCGG CCTTggcTG  
 451 AAAGAAAAAa ccAGCATTAT CAATGTGcGc GGAATGTTGC CCGACCATAC  
 501 GCTTTTGGGC ATCAAAATTT GGGCGCGCAA CGATAAAAC GAATTGGCAG  
 551 AGGCAGTGGA AGCCGATTCC GCCGTTTGA ACAGCGACGG CAGCTGGCAG  
 601 TTGAAAAACA TCCGCGCAG CATCATGGGT ACAGACAAA TCGAAACATC  
 35 651 cgCGCGCC GAAGAACTT gGCCGATTGC CGTCAGACGC AACCTGATGG  
 701 ACGTATTGCT CGTCAAGCCC GACCAAATGT CCGTCGGCGA GCTGACCACC  
 751 TACATCCGCC ACCTCCAAA CAACAGCCAA AACACCCAAA TCTACGCCAT  
 801 CGCATGGTGG CGTAAACTCG TTACCCCGT CGCCGCATGG GTCATGGCGC  
 851 TCGTTGCCTT CGCCTTACG CCGCAAACCA CGCGCCACGG CAATATGGGC  
 40 901 TTAATACTCT TCGGCGCAT CTGTCTCGGA TTGCTGTTCC ACCTTGGCCG  
 951 CAGGCTCTT GGGTTTACCA GCCAACTCTA CGGCACCCA CCCTTCCTCG  
 1001 CCGGCGCACT GCCTACCATA GCCTTCGCCT TGCTCGCTGT TTGGCTGATA  
 1051 CGCAAACAGG AAAACGTTG A

This encodes a protein having amino acid sequence <SEQ ID 892>:

45 1 MNLISRYIIR QMAVMNAVYAL LAFLALYSFF EILYETGNLG KGSYGIWEML  
 51 GYTALKMPAR AYELMPLAVL IGGLASLSQL AAGSELAVIK ASGMSTKKLL  
 101 LILSQFGFIF AIAAVALGEW VAPTLSQKAE NIKAAAINGK ISTGNTGLWL  
 151 KEKTSIINVR GMLPDHTLLG IKIWARNDKN ELAEAVEADS AVLNSDGSWQ  
 201 LKNIRRSIMG TDKIETSAAS EETWPIAVRR NLMDVLLVKP DQMSVGELTT  
 50 251 YIRHLQNNNSQ NTQIYIAWW RKLVPVAAW VMALVAFATP QTTRHGNMG  
 301 LKLFGGICLG LLFHLAGRLF GFTS QLYGTP PFLAGALPTI AFALLAVWLI  
 351 RKQEK\*

ORF112ng and ORF112-1 show 94.2% identity in 326 aa overlap:

55 orf112ng MNLISRYIIRQMAVMNAVYALLAFLALYSFFEILYETGNLGKGSYGIWEMLGYTALKMPAR  
 orf112-1 MNLISRYIIRQMAVMNAVYALLAFLALYSFFEILYETGNLGKGSYGIWEMLGYTALKMPAR  
 60 10 20 30 40 50 60  
 70 80 90 100 110 120

5	orf112ng	AYELMPLAVLIGGLASLSQLAAGSELAVIKASGMSTKKLLLLLSQFGFIFAIAAVALGEW
	orf112-1	AYELIPLAVLIGGLVLSLSQLAAGSELTVIKASGMSTKKLLLLLSQFGFIFAIATVALGEW
10	orf112ng	VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKTSIINVRGMLPDHTLLGIKIWARNDKN
	orf112-1	VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKSNXINVREMLPDHTLLGIKIWARNDKN
15	orf112ng	ELAEAVEADSAVLNSDGSWQLKNIRRSIMGTDKIETSAAAEETWPIAVRRNLMDVLLVKP
	orf112-1	ELAEAVEADSAVLNSDGSWQLKNIRRTLGEDKVEVSIAAEENWPISVKRNLMDVLLVKP
20	orf112ng	DQMSVGELTTYIRHLQNNSQNTQIYAIAWWRKLVYPVAAWVMALVAFATPQTTRHGNMG
	orf112-1	DQMSVGELTTYIRHLQNNSQNTRIYAIAWWRKLVYPAAWVMALVAFATPQTTRHGNMG
25	orf112ng	LKLFGGICLGLLFHLAGRLFGFTSQLYGTTPFLAGALPTIAFALLAVWLIRKQEKRX
	orf112-1	LKLFGGICXGLLFHLAGRLFGFTSQL

- 30 This analysis suggests that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 105

Table III lists several *Neisseria* strains which were used to assess the conservation of the sequence of ORF 4 among different strains.

35 **TABLE III – List of *Neisseria* Strains Used for Gene Variability Study of ORF 4**

#### ORF4 gene variability: List of used *Neisseria* strains

Identification Strains number		Source / reference
<b>Group B</b>		
zv01_4	NG6/88	R. Moxon / Seiler <i>et al.</i> , 1996
zv02_4	BZ198	R. Moxon / Seiler <i>et al.</i> , 1996
zv03_4ass	NG3/88	R. Moxon / Seiler <i>et al.</i> , 1996
zv04_4	297-0	R. Moxon / Seiler <i>et al.</i> , 1996
zv05_4	1000	R. Moxon / Seiler <i>et al.</i> , 1996
zv06_4	BZ147	R. Moxon / Seiler <i>et al.</i> , 1996
zv07_4	BZ169	R. Moxon / Seiler <i>et al.</i> , 1996
zv08_4	528	R. Moxon / Seiler <i>et al.</i> , 1996
zv09_4	NGP165	R. Moxon / Seiler <i>et al.</i> , 1996
zv10_4	BZ133	R. Moxon / Seiler <i>et al.</i> , 1996

zv11_4	NGE31	R. Moxon / Seiler <i>et al.</i> , 1996
zv12_4ass	NGF26	R. Moxon / Seiler <i>et al.</i> , 1996
zv13_4	NGE28	R. Moxon / Seiler <i>et al.</i> , 1996
zv15_4	SWZ107	R. Moxon / Seiler <i>et al.</i> , 1996
zv16_4	NGH15	R. Moxon / Seiler <i>et al.</i> , 1996
zv17_4	NGH36	R. Moxon / Seiler <i>et al.</i> , 1996
zv18_4	BZ232	R. Moxon / Seiler <i>et al.</i> , 1996
zv19_4	BZ83	R. Moxon / Seiler <i>et al.</i> , 1996
zv20_4	44/76	R. Moxon / Seiler <i>et al.</i> , 1996
zv21_4	MC58	R. Moxon
zv96_4	2996	Our collection
<b>Group A</b>		
zv22_4	205900	R. Moxon
z2491_4	Z2491	R. Moxon / Maiden <i>et al.</i> , 1998
<b>Group C</b>		
zv24_4	90/18311	R. Moxon
zv25_4	93/4286	R. Moxon
<b>Others</b>		
zv26_4ass	A22 (group W)	R. Moxon / Maiden <i>et al.</i> , 1998
zv27_4	E26 (group X)	R. Moxon / Maiden <i>et al.</i> , 1998
zv28_4	860800 (group Y)	R. Moxon / Maiden <i>et al.</i> , 1998
zv29_4	E32 (group Z)	R. Moxon / Maiden <i>et al.</i> , 1998
<b>Gonococcus</b>		
zv32_4	Ng F62	R. Moxon / Maiden <i>et al.</i> , 1998
zv33_4	Ng SN4	R. Moxon
fa1090_4	FA1090	R. Moxon

**References:**

Seiler A. *et al.*, Mol. Microbiol., 1996, 19(4):841-856.  
Maiden *et al.*, Proc. Natl. Acad. Sci. USA, 1998, 95:3140-3145.

The amino acid sequences for each listed strain are as follows:

```
>FA1090_4 <SEQ ID 893>
MKTFFKTL$AAALALILAACGGQKDSAPAASAAAPSADNGAAKKEIVFGTTVGDFGDMVK
EQIQAELEKKGYTVKLVEFTDYVRPNLALAEGELDINVQHKPYLDDFKKEHNL$DITEAF
QVPTAPLGLYPGKLK$LEE$VKD$G$TV$APND$PSNFARALVMLNELGWIKLKDGINPLTAS
```

KADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNW  
SAVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNEGAAK\*

5 >Z2491\_4 <SEQ ID 894>  
MKTFFKTLAALALILAACGGQKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDMVKE  
QIQAELEKKGYTVKLVEFTDYVRPNLALAEGELDINVQHKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLKSLEEVKDGSTVSAPNDPSNFARVLVMLDELGWIKLKDGINPLTASK  
10 ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNW  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNEGAAK\*

>ZV01\_4 <SEQ ID 895>  
MKTFFKTLAALALILAACGGQKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDMVKE  
QIQAELEKKGYTVKLVEFTDYVRPNLALAEGELDINVQHKPYLDDFKKEHNLDITEVFQ  
15 VPTAPLGLYPGKLKSLEEVKDGSTVSAPNDPSNFARVLVMLDELGWIKLKDGINPLTASK  
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNW  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNEGAAK\*

>ZV02\_4 <SEQ ID 896>  
MKTFFKTLAALALILAACGGQKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDMVKE  
20 HIQPELEKKGYTVKLVEFTDYVRPNLALAEGELDINVQHKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLKSLEEVKDGSTVSAPNDPSNFARVLVMLDELGWIKLKDGINPLTASK  
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNW  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNEGAAK\*

25 >ZV03\_4ASS <SEQ ID 897>  
MKTFFKTLAALALILAACGGQKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDMVKE  
HIQPELEKKGYTVKLVEFTDYVRPNLALAEGELDINVQHKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLKSLEEVKDGSTVSAPNDPSNFARVLVMLDELGWIKLKDGINPLTASK  
30 ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNW  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNEGAAK\*

>ZV04\_4 <SEQ ID 898>  
MKTFFKTLAALALILAACGGQKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDMVKE  
35 HIQPELEKKGYTVKLVEFTDYVRPNLALAEGELDINVQHKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLKSLEEVKDGSTVSAPNDPSNFARVLVMLDELGWIKLKDGINPLTASK  
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNW  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNEGAAK\*

40 >ZV05\_4 <SEQ ID 899>  
MKTFFKTLAALALILAACGGQKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDMVKE  
HIQPELEKKGYTVKLVEFTDYVRPNLALAEGELDINVQHKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLKSLEEVKDGSTVSAPNDPSNFARVLVMLDELGWIKLKDGINPLTASK  
45 ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNW  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNEGAAK\*

>ZV06\_4 <SEQ ID 900>  
MKTFFKTLAALALILAACGGQKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDMVKE  
50 QIQAELEKKGYTVKLVEFTDYVRPNLALAEGELDINVQHKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLKSLEEVKDGSTVSAPNDPSNFARVLVMLDELGWIKLKDGINPLTASK  
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNW  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNEGAAK\*

55 >ZV07\_4 <SEQ ID 901>  
MKTFFKTLAALALILAACGGQKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDMVKE  
QIQAELEKKGYTVKLVEFTDYVRPNLALAEGELDINVQHKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLKSLEEVKDGSTVSAPNDPSNFARVLVMLDELGWIKLKDGINPLTASK  
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNW  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNEGAAK\*

60 >ZV08\_4 <SEQ ID 902>  
MKTFFKTLAALALILAACGGQKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDMVKE  
HIQPELEKKGYTVKLVEFTDYVRPNLALAEGELDINVQHKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLKSLEEVKDGSTVSAPNDPSNFARVLVMLDELGWIKLKDGINPLTASK  
65 ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNW  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNEGAAK\*

>ZV09\_4 <SEQ ID 902>  
MKTFFKTLAALALILAACGGQKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDMVKE  
70 HIQPELEKKGYTVKLVEFTDYVRPNLALAEGELDINVQHKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLKSLEEVKDGSTVSAPNDPSNFARVLVMLDELGWIKLKDGINPLTASK  
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNW  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNEGAAK\*

75 >ZV10\_4 <SEQ ID 903>  
MKTFFKTLAALALILAACGGQKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDMVKE  
HIQPELEKKGYTVKLVEFTDYVRPNLALAEGELDINVQHKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLKSLEEVKDGSTVSAPNDPSNFARVLVMLDELGWIKLKDGINPLTASK  
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNW  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNEGAAK\*

>ZV11\_4 <SEQ ID 904>  
MKTFFKTLAALALILAAACGGQKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDMVKE  
QIQAELEKKGYTVKLVEFTDYVRPNLALAEGLDINVFOHKKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDLDELGWIKLKDGINPLTASK  
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEFGYKSPAANWEGAAG\*

>ZV12\_4ASS <SEQ ID 905>  
MKTFFKTLAALALILAAACGGQKDRAPAASASAAASENGAAKKEILFGTTVGDLGDMVKE  
QIQAELEKKGYTVKLVEFTDYVRPNLALAEGLDINVFOHKKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDLDELGWIKLKDGINPLTASK  
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEFGYKSPAANWEGAAG\*

>ZV13\_4 <SEQ ID 906>  
MKTFFKTLAALALILAAACGGQKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDMVKE  
QIQAELEKKGYTVKLVEFTDYVRPNLALAEGLDINVFOHKKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDLDELGWIKLKDGINPLTASK  
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEFGYKSPAANWEGAAG\*

>ZV15\_4 <SEQ ID 907>  
MKTFFKTLAALALILAAACGGQKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDMVKE  
HIQAELEKKGYTVKLVEFTDYVRPNLALAEGLDINVFOHKKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDLDELGWIKLKDGINPLTASK  
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEFGYKSPAAGNEGAAG\*

>ZV16\_4 <SEQ ID 908>  
MKTFFKTLAALALILAAACGGQKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDMVKE  
HIQAELEKKGYTVKLVEFTDYVRPNLALAEGLDINVFOHKKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDLDELGWIKLKDGINPLTASK  
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEFGYKSPAANWEGAAG\*

>ZV17\_4 <SEQ ID 909>  
MKTFFKTLAALALILAAACGGQKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDMVKE  
QIQAELEKKGYTVKLVEFTDYVRPNLALAEGLDINVFOHKKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDLDELGWIKLKDGINPLTASK  
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEFGYKSPAANWEGAAG\*

>ZV18\_4 <SEQ ID 910>  
MKTFFKTLAALALILAAACGGQKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDMVKE  
HIQAELEKKGYTVKLVEFTDYVRPNLALAEGLDINVFOHKKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDLDELGWIKLKDGINPLTASK  
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEFGYKSPAANWEGAAG\*

>ZV19\_4 <SEQ ID 911>  
MKTFFKTLAALALILAAACGGQKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDMVKE  
QIQAELEKKGYTVKLVEFTDYVRPNLALAEGLDINVFOHKKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDLDELGWIKLKDGINPLTASK  
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEFGYKSPAANWEGAAG\*

>ZV20\_4 <SEQ ID 912>  
MKTFFKTLAALALILAAACGGQKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDMVKE  
QIQAELEKKGYTVKLVEFTDYVRPNLALAEGLDINVFOHKKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDLDELGWIKLKDGINPLTASK  
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEFGYKSPAANWEGAAG\*

>ZV21\_4 <SEQ ID 913>  
MKTFFKTLAALALILAAACGGQKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDMVKE  
QIQAELEKKGYTVKLVEFTDYVRPNLALAEGLDINVFOHKKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDLDELGWIKLKDGINPLTASK  
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEFGYKSPAANWEGAAG\*

>ZV22\_4 <SEQ ID 914>  
MKTFFKTLAALALILAAACGGQKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDLVKE  
QIQAELEKKGYTVKLVEFTDYVRPNLALAEGLDINVFOHKKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDLDELGWIKLKDGINPLTASK  
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEFGYKSPAANWEGAAG\*

>ZV24\_4ASS <SEQ ID 915>

MKTFFKTL SAAALALILAACGGQKDSAPAASASAAAADNGAEKKEIVFGTTVGDFGDMVKE  
 HIQPELEKKGYTVKLVEFTDDVVRPNLALAE GELDINV FQH KPYLDDFKKEHNL DITEVFQ  
 VPTAPLGLYPGKLKSL EEVKDGSTVSAPNDPSNFARVLV MLDELGWIKLKDGINPLTASK  
 ADIAENLNKIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS  
 AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFE GYKSPAAWNEGA AK\*

>ZV25\_4 <SEQ ID 916>

MKTFFKTL SAAALALILAACGGQKDSAPAASASAAAADNGAEKKEIVFGTTVGDFGDMVKE  
 HIQPELEKKGYTVKLVEFTDYVRPNLALAE GELDINV FQH KPYLDDFKKEHNL DITEVFQ  
 VPTAPLGLYPGKLKSL EEVKDGSTVSAPNDPSNFARVLV MLDELGWIKLKDGINPLTASK  
 ADIAENLNKIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS  
 AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFE GYKSPAAWNEGA AK\*

>ZV26\_4 <SEQ ID 917>

MKTFFKTL SAAALALILAACGGQKDSAPAASASAAAADNGAEKKEIVFGTTVGDFGDMVKE  
 HIQPELEKKGYTVKLVEFTDYVRPNLALAE GELDINV FQH KPYLDDFKKEHNL DITEVFQ  
 VPTAPLGLYPGKLKSL EEVKDGSTVSAPNDPSNFARVLV MLDELGWIKLKDGINPLTASK  
 ADIAENLNKIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS  
 AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFE GYKSPAAWNEGA AK\*

>ZV27\_4 <SEQ ID 918>

MKTFFKTL SAAALALILAACGGQKDSAPAASASAAAADNGAEKKEIVFGTTVGDFGDMVKE  
 HIQPELEKKGYTVKLVEFTDYVRPNLALAE GELDINV FQH KPYLDDFKKEHNL DITEVFQ  
 VPTAPLGLYPGKLKSL EEVKDGSTVSAPNDPSNFARVLV MLDELGWIKLKDGINPLTASK  
 ADIAENLNKIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS  
 AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFE GYKSPAAWNEGA AK\*

>ZV28\_4 <SEQ ID 919>

MKTFFKTL SAAALALILAACGGQKDSAPAASASAAAADNGAEKKEIVFGTTVGDFGDMVKE  
 HIQPELEKKGYTVKLVEFTDYVRPNLALAE GELDINV FQH KPYLDDFKKEHNL DITEVFQ  
 VPTAPLGLYPGKLKSL EEVKDGSTVSAPNDPSNFARVLV MLDELGWIKLKDGINPLTASK  
 ADIAENLNKIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS  
 AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFE GYKSPAAWNEGA AK\*

>ZV29\_4 <SEQ ID 920>

MKTFFKTL SAAALALILAACGGQKDSAPAASASAAAADNGAEKKEIVFGTTVGDFGDMVKE  
 HIQPELEKKGYTVKLVEFTDYVRPNLALAE GELDINV FQH KPYLDDFKKEHNL DITEVFQ  
 VPTAPLGLYPGKLKSL EEVKDGSTVSAPNDPSNFARVLV MLDELGWIKLKDGINPLTASK  
 ADIAENLNKIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS  
 AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFE GYKSPAAWNEGA AK\*

>ZV32\_4 <SEQ ID 921>

MKTFFKTL SAAALALILAACGGQKDSAPAASAAAADNGAEKKEIVFGTTVGDFGDMVKE  
 EQIQAELEKKGYTVKLVEFTDYVRPNLALAE GELDINV FQH KPYLDDFKKEHNL DITEAF  
 QVPTAPLGLYPGKLKSL EEVKDGSTVSAPNDPSNFARVLV MLDELGWIKLKDGINPLTASK  
 KADIAENLNKIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS  
 SAVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFE GYKSPAAWNEGA AK\*

>ZV33\_4 <SEQ ID 922>

MKTFFKTL SAAALALILAACGGQKDSAPAASAAAADNGAEKKEIVFGTTVGDFGDMVKE  
 EQIQAELEKKGYTVKLVEFTDYVRPNLALAE GELDINV FQH KPYLDDFKKEHNL DITEAF  
 QVPTAPLGLYPGKLKSL EEVKDGSTVSAPNDPSNFARVLV MLDELGWIKLKDGINPLTASK  
 KADIAENLNKIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS  
 SAVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFE GYKSPAAWNEGA AK\*

>ZV96\_4 <SEQ ID 923>

MKTFFKTL SAAALALILAACGGQKDSAPAASASAAAADNGAEKKEIVFGTTVGDFGDMVKE  
 HIQAELEKKGYTVKLVEFTDYVRPNLALAE GELDINV FQH KPYLDDFKKEHNL DITEVFQ  
 VPTAPLGLYPGKLKSL EEVKDGSTVSAPNDPSNFARVLV MLDELGWIKLKDGINPLTASK  
 ADIAENLNKIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS  
 AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFE GYKSPAAWNEGA AK\*

Figure 8 shows the results of aligning the sequences of each of these strains. Dark shading indicates regions of homology, and gray shading indicates the conservation of amino acids with similar characteristics. As is readily discernible, there is significant conservation among the various strains of ORF 4, further confirming its utility as an antigen for both vaccines and diagnostics.